PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/54, 15/11, 15/82, 9/10, A01H 5/00, 5/10 — A2

(11) International Publication Number: **WO 00/70058**

(43) International Publication Date: 23 November 2000 (23.11.00)

(21) International Application Number: PCT/GB00/01890

(22) International Filing Date: 17 May 2000 (17.05.00)

(30) Priority Data:
9911379.7 18 May 1999 (18.05.99) GB

(71) Applicant *(for all designated States except US)*: THE VICTORIA UNIVERSITY OF MANCHESTER [GB/GB]; Oxford Road, Manchester M13 9PL (GB).

(72) Inventors; and

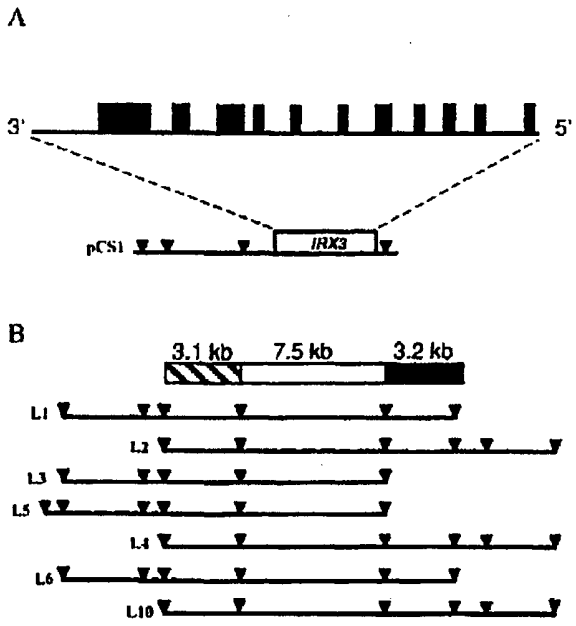
(75) Inventors/Applicants *(for US only)*: TURNER, Simon [GB/GB]; 17 Knight Street, Didsbury, Manchester M20 6WG (GB). TAYLOR, Neil [GB/GB]; 84 Balbrook Court, Wilmslow Road, Didsbury, Manchester M20 3QT (GB).

(74) Agent: STARK, Amanda, Jane; Marks & Clerk, Sussex House, 83–85 Mosley Street, Manchester M2 3LG (GB).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PLANT CELLULOSE SYNTHASE GENES



(57) Abstract

The present invention relates to an isolated nucleic acid molecule comprising a cellulose synthase gene specifically expressed during deposition of secondary cell walls in lignin containing cells and the use of such a gene or its promoter to modulate the expression of enzymes involved in the synthesis of plant cell walls, to produce transgenic plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

PLANT CELLULOSE SYNTHASE GENES

The present invention relates to plant cellulose synthase genes and their use in modifying plant phenotypes.

The invention also relates to constructs containing the cellulose synthase gene or a promoter thereof and the use of such constructs to regulate the expression of genes specifically during secondary cell wall deposition in lignin containing cells.

Cellulose forms the structural framework of plant cell walls and is probably the world's most abundant biopolymer. Cellulose is made up of crystalline β-1,4-glucan microfibrils. These crystalline microfibrils are extremely strong and resist enzymic and mechanical degradation. For many plant cells, the cell wall is synthesised in two distinct stages. During the initial phase of cellular growth, a primary cell wall is laid down and continuously expanded by processes that include relaxation of interchain linkages and addition of new polymers and matrix materials. Cellulose usually comprises about 20 to 30% of the dry weight of the primary wall (Fry, 1988). Following the cessation of expansion and division, a secondary cell wall is synthesised within the bounds of the primary wall. Cellulose accounts for roughly 40 to 90% of the secondary cell wall, depending upon the cell type.

The deposition of secondary wall material often results in a very thick wall and is responsible for many of the structural properties associated with plants. In some heavily thickened cells, such as xylem cells, the secondary wall may also contain a high proportion of lignin that contributes to the mechanical strength. Consequently, the many industrial processes that utilise plant material, which are as diverse as paper manufacturing and food processing depend heavily on the properties of plant secondary cell walls. It would therefore be advantageous to modify the structure and cellulose content of plant secondary cell walls to produce altered plant phenotypes specific to the needs of a particular industry, for example reducing the lignin content of wood pulp for paper manufacturing.

The mechanisms involved in the synthesis of secondary cell walls are not understood in detail (Emons and Mulder, 1998). It is generally accepted that the cellulose component of both primary and secondary cell walls is synthesised by enzyme complexes situated at the plasma membrane. Many freeze-fracture studies have identified plasma membrane particles known as rosettes that appear to be associated with the ends of microfibrils (Brown 1996).

The spacing of these rosettes also correlates with the distribution of the microfibrils (Giddings *et al.*, 1980). It has been suggested that each rosette consists of a hexameric complex, which result in the synthesis of 36 β-glucan chains that are thought to be present in a primary microfibril (Delmer and Amor, 1995). The differences in physical properties of primary and secondary plant cell walls are partly due to differences in the number of individual cellulose chains in the microfibril unit. In contrast to the approximately 36 individual chains in primary microfibrils (Delmer and Amor, 1995), the secondary cell walls of some algae contain fibrils containing up to 12000 individual β-1,4-glucan chains (Brown *et al.*, 1996). In addition, individual cellulose chains from the secondary wall typically contain about 14,000 β-1,4-linked glucose molecules, whereas in the primary wall about half of the cellulose molecules contain less than about 500 glucose moieties and half contain about 2500-4500 monomers (Blaschek *et al.*, 1982).

The enzyme complex which catalyses the synthesis of cellulose in plants is termed cellulose synthase. Cellulose synthase from higher plants is assumed to be a multi-enzyme complex (Delmer and Amor, 1995). Consistent with this concept, a four-gene operon responsible for cellulose synthesis has been cloned from *Acetobacter xylinum* (Saxena *et al.*, 1990), and five genes have been shown to be essential for cellulose synthesis in Agrobacterium (Matthese *et al.*, 1995). Only one of these genes shows sequence similarity between Agrobacterium and *A. xylinum* and this gene has been identified as encoding the cellulose synthase catalytic subunit. Amino acid sequences of bacterial cellulose synthases along with other enzymes requiring nucleotide sugars were found to contain four regions of high conservation thought to be critical for UDP-glucose binding and catalysis (Saxena *et al.*, 1995).

Recently, cDNA clones for two cellulose synthase homologues containing all four conserved regions were identified from a cotton cDNA library prepared from fibres at the onset of secondary cell wall synthesis (Pear *et al.*, 1996). These genes, which are termed *CELA* genes, exhibit sequence similarity to at least 31 distinct expressed sequence tag (EST) or genomic sequences in the Arabidopsis sequence databases (Cutler and Somerville, 1997). However, it is unlikely that all of these cellulose synthase-like (*CSL*) genes actually catalyse cellulose synthesis (Cutler and Somerville, 1997; Delmer, 1998). Rather, it has been proposed that some of the *CSL* genes encode other glycan synthases, such as those responsible for the synthesis of xyloglucan, xylan, callose and other polysaccharides.

The biological function of one of the *CELA*-related genes was recently established by the characterisation of a mutant of Arabidopsis deficient in crystalline cellulose deposition. The *radial swelling 1* (*rsw1*) mutant exhibits temperature sensitive radial swelling of its root tip due to a deficiency in cellulose deposition at elevated temperature (Baskin *et al.*, 1992). The *RSW1* gene encodes a polypeptide with a high degree of sequence similarity to the cotton *CELA* genes (Arioli *et al.*, 1998a). The *RWS1* gene appears to affect cellulose synthesis in primary cell walls, in that plants with the *rws1* mutation are not viable and do not grow past the seedling stage.

International patent application number PCT/US97/19529 to Calgene states that one of the cotton fibre *CELA* genes, *CELA1* is expressed in developing cotton fibres when secondary cell wall synthesis is initiated. The application shows how the CELA genes were used to screen the dBEST databank of rice and Arabidopsis ESTs to identify cDNA clones with homologous sequences from these plants. There is no teaching that any of these homologous sequences encode a protein having cellulose synthase activity or that any of the homologous genes are expressed at a particular time during plant development or in specific tissues.

PCT/US97/19529 describes how the cotton fibre CELA1 promoter may be used in a promoter construct and postulates that the constructs may be used in conjunction with plant regeneration systems to obtain plant cells and plants, and allow the phenotype of fibre cells to be modified to provide cotton fibres which are coloured as a result of genetic engineering. PCT/US97/19529 further postulates that the gene described therein may be used in a construct to transform woody tissues so that they produce excess cellulose, thereby reducing lignin production.

There is no disclosure in PCT/US97/19529 as to what construct would be used to transform forest tree species so as to modify the wood quality phenotype, and to suppress lignin production. As the secondary cell wall of a developing cotton fibre is almost pure cellulose and does not contain lignin it would appear unlikely that the CELA1 gene would be expressed in woody tissue and thus its promoter would not be expected to be useful in a construct for transforming forest tree species.

For many applications it is desirable to be able to control gene expression at a particular stage in the growth of a plant or in a particular tissue. For this purpose regulatory sequences are required to turn on transcription at a particular time in a plant's development or in a

particular tissue without effecting expression of other genes. As it is the composition of secondary cell walls that is generally important for the paper, pulp and food processing industries it is desirable to provide a gene which affects synthesis of cell wall components specifically in secondary cell walls of woody plants. Furthermore it is desirable to provider control of expression of genes during secondary cell wall deposition so as to be able to alter the phenotype of woody plants.

Accordingly, the first aspect of the invention provides an isolated nucleic acid molecule comprising a cellulose synthase gene specifically expressed during deposition of secondary cell walls in lignin containing cells.

The invention is based on the inventors' work on mutants of Arabidopsis carrying mutations in one of the three *irx* (for *irregular xylem*) loci. These genes are characterised by collapsed xylem in stems (Turner and Somerville 1997). The xylem vessels are thought to collapse due to a lack of resistance to the negative pressure exerted by water transport. The deposition of cell walls in these plants is abnormal and results in the stems being weaker and less rigid. In one of these mutants, *irx3*, the increased flexibility of the stems results in an inability to support an upright growth habit. Analysis of these mutants showed a specific reduction or complete loss of cellulose deposition in the secondary cell wall (Turner and Somerville, 1997).

The inventors have isolated and characterised of a member of the Arabidopsis *CELA* gene family that corresponds to the *IRX3* gene. The discovery that *IRX3* is a component of the cellulose synthases involved in secondary wall synthesis created several experimental opportunities for studies of the factors that regulate secondary wall synthesis and lead to the present invention.

Preferably, the cellulose synthase gene according to the first aspect of the invention is specifically expressed during deposition of secondary cell walls in vascular tissue such as xylem. This is evidenced by the collapsed xylem in *irx3* mutants which do not express the *IRX3* gene.

The preferred cellulose synthase gene is that isolated from Arabidopsis. The preferred sequence of the cellulose synthase gene according to the first aspect of the invention is that comprising the sequence shown as SEQ ID No. 1, the complement of the sequence shown as

SEQ ID No. 1, the reverse complement of the sequence shown as SEQ ID No. 1, the reverse of the sequence shown as SEQ ID No. 1 or a sequence having at least 80 % sequence identity with the nucleic acid molecule sequences of any one of the aforementioned sequences.

By use of the term "at least 80% identity" it is therefore understood that the invention also encompasses more than the specific exemplary nucleotide sequences. Modifications to the sequence, such as deletions, insertions, or substitutions in the sequence which produce "silent" changes which do not substantially affect the functional properties of the resulting protein molecule are also contemplated. For example, alterations in the nucleotide sequence which reflect the degeneracy of the genetic code or which result in the production of a chemically equivalent amino acid at a given site are contemplated.

Nucleotide changes which result in an alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein.

A nucleic acid sequence with a greater identity than 80 % to SEQ 1D No. 1 is also envisaged. Preferably, the nucleic acid sequence has 85 % identity with SEQ ID No.1, more preferably 90 % identity, even more preferably 95 % identity and most preferably 98% identity with SEQ ID No. 1.

The cellulose synthase gene according to the first aspect of the invention comprises the cellulose synthase promoter and the cellulose synthase coding region. The promoter is time and tissue specific in that it turns on expression of the cellulose synthase gene only during secondary cell wall synthesis and only in cells containing lignin, such as vascular tissue. The promoter thus provides an important second aspect of the invention.

According to a second aspect of the invention an isolated nucleic acid molecule containing a promoter of an isolated nucleic acid molecule comprising a cellulose synthase gene specifically expressed during deposition of secondary cell walls in lignin containing cells is provided.

Preferably, the cellulose synthase promoter regulates expression of the cellulose synthase gene so that it is expressed only during deposition of secondary cell walls in vascular tissue such as xylem.

As with the cellulose synthase gene described in accordance with the first aspect of the invention the preferred cellulose synthase promoter is that isolated from Arabidopsis. The preferred sequence of the cellulose synthase promoter according to the second aspect of the invention is that comprising the sequence shown as SEQ ID No. 3 or SEQ ID NO 4, the complement of the sequence shown as SEQ ID No. 3 or SEQ ID NO 4, the reverse complement of the sequence shown as SEQ ID No. 3 or SEQ ID NO 4, the reverse of the sequence shown as SEQ ID No. 3 or SEQ ID NO 4 or a sequence having at least 60 % sequence identity with the nucleic acid molecule sequences of any one of the aforementioned sequences.

As with the gene sequence, base changes may be present in a promoter sequence without substantially affecting its functionality. Such modifications are within the scope of the invention.

A nucleic acid sequence with a greater identity than 60 % to SEQ 1D No. 3 or SEQ ID NO 4 is also envisaged. Preferably, the nucleic acid sequence has 70 % identity with SEQ ID No.3 or 4, more preferably 80 % identity, even more preferably 90 % identity and most preferably 95% identity with SEQ ID No. 3 or SEQ ID NO 4.

Suitable nucleic acid sequences selected according to the invention may be obtained, for example, by cloning techniques using cDNA libraries corresponding to a wide variety of plant species expressing lignin. Suitable nucleotide sequences may be isolated from DNA libraries obtained from a wide variety of species by means of nucleic acid hybridisation or PCR, using as hybridisation probes or primers nucleotide sequences selected in accordance with the invention, such as SEQ ID No 1 or SEQ ID NO 3 or specific fragments thereof.

Since the promoter according to the second aspect of the invention is both developmentally and tissue specific it may advantageously be linked to an exogenous gene and used to transform a plant, such that that gene is only expressed in the transformed plant during secondary cell wall synthesis and only in tissues containing lignin.

According to the third aspect of the invention there is provided a nucleic acid construct suitable for transforming a plant cell, the construct comprising, in the 5'-3' direction:

(a) a cellulose synthase promoter according to the second aspect of the invention, and

(b) a nucleotide sequence of an exogenous gene;

the construct being arranged such that expression of the exogenous gene is under the control of the promoter.

The constructs may be used to provide for transcription of a nucleotide sequence of interest in cells of a plant host that produces lignin, only during secondary cell wall synthesis. The constructs may take several forms depending on the intended use of the construct. The constructs include vectors, transcriptional cassettes, plasmids and expression cassettes.

In one embodiment the nucleic acid construct includes a coding sequence for at least a functional part of an enzyme involved in synthesis of plant cell wall components. Generally, the enzyme may be involved in synthesis of cell wall polysaccharide biosynthesis or cell wall protein biosynthesis. More particularly it is preferred that the construct comprises a nucleotide sequence encoding at least a functional part of an enzyme involved in cellulose biosynthesis or lignin biosynthesis.

For applications where amplification of a particular protein is desired, the nucleotide sequence is inserted in the construct in a sense orientation, such that transformation of the target plant with the construct will lead to an increase in the number of copies of the gene and therefore an increase in an amount of enzyme.

When down regulation of a particular protein is desired the nucleotide sequence is inserted in the construct in an antisense orientation such that RNA produced by the transcription of the nucleotide sequence is complementary to the endogenous mRNA sequence. This, in turn, will result in a decrease in the number of copies of the gene and therefore a decrease in the amount of enzyme.

As an alternative the nucleic acid construct may comprise a nucleotide sequence including a non-coding region of an exogenous gene or a sequence complementary to such a sequence. As used here the term "non-coding region" includes both transcribed sequences which are not translated and non-transcribed sequences within about 1000 base pairs 5' or 3' of the translated sequences or open reading frames. Examples of non-coding regions which could be useful according to the third aspect of the invention include introns and 5' non-coding leader sequences. Transformation of a target plant with such a DNA construct may lead to

the reduction in the amount of a particular protein or polysaccharide synthesised by the plant by the process of co-suppression.

According to a preferred embodiment the construct comprises the antisense of nucleotide sequence encoding an enzyme involved in lignin biosynthesis.

The constructs of the present invention may be used to transform a variety of plants, both monocotyledonous (e.g. corn, grains, grasses, oil seed rape, barley, rice, forage grasses, wheat and oat), dicotyledonous (e.g. Arabidopsis, tobacco, legumes, alfalfa, oaks, maple, poplar and eucalyptus) and gymnosperms (e.g. Scots pine, white spruce and larch). In a preferred embodiment the constructs are used to transform woody plants, herein defined as a tree or shrub whose stem lives for a number of years and increases in diameter each year by the addition of woody tissue.

Techniques for stably incorporating the constructs into the genome of target plants are well known in the art and include Agrobacterium tumefaciens mediated introduction, electroporation, protoplast fusion, injection into reproductive organs, high velocity projectile introduction an similar methods.

Transformed transgenic plant cells are then placed in an appropriate selective medium for selection of transgenic cells which are then grown to callus, shoots grown and plantlets generated from the shoot by growing in rooting media.

To confirm the presence of transformed cells a Southern blot analysis may be performed using methods familiar to those skilled in the art. The plants may be harvested and/or the seeds collected. The seed may serve as a source for growing additional plants having the desired characteristics.

Of particular importance in the use of the constructs according to the third aspect of the invention is the ability to obtain plants whose phenotype is altered in a tissue specific and developmentally specific manner. By using the cellulose synthase gene which is only expressed during secondary cell wall synthesis and only in cells containing lignin or vascular tissue it is possible to produce a plant which is normal during it primary growth phase and only exhibits and altered phenotype during the secondary growth phase.

A particularly preferred method of use of the construct is to reduce the amount of lignin in woody tissues, although the principle is equally applicable to other secondary cell wall components.

Lignin is a major problem for the pulp and paper industry and considerable effort is used in removing lignin from paper pulp. Many groups have used an antisense approach, which involves expressing a lignin biosynthesis gene in reverse orientation and expressing it in cells making lignin (i.e. secondary cell walls in some plants) in order to reduce the lignin content of trees. In order to express these antisense genes, the correct promoter is required to direct expression in secondary cell walls. To date the promoters of lignin biosynthesis genes or other promoters have been used. The promoter described according to the second aspect of the invention may be useful for such a purpose. It is postulated that because the cellulose synthase promoter may be activated before the lignin biosynthesis genes that it may be a better promoter than those known in the art for altering lignin in secondary cell walls.

The invention will now be described, by way of example only, with reference to the following figures, in which:

Figure 1 illustrates the localisation of the *irx3* mutation on chromosome V.
The positions of YAC clones spanning this region are shown below (from Schmidt *et al.*, 1996). The YAC clones containing the *IRX3* gene filled. The filled vertical bar indicates the region of the chromosome V containing the *IRX3* gene. The positions of genetic markers are taken from the map generated from recombinant inbred lines (Lister and Dean 1993).

Figure 2 illustates a map of genomic clones containing the *IRX3* gene.
Introns are represented by solid blocks and triangles indicate the position of HindIII sites. Boxes represent the positions of the 3.1 kb (hatched), 7.5 kb (open), and 3.2 kb (filled) HindIII fragments referred to in the text. Two additional HindIII sites not shown occur between the 7.5 kb and 3.2 kb HindIII fragments.
(A) clone used to subclone the *IRX3* gene and intron/exon map of the *IRX3* gene.
(B) Cosmid clones used for complementation.

Figure 3 illustrates alignment of the amino acid sequences of plant cellulose synthase genes. Solid boxes indicate regions in which more that half the residues are identical, and grey boxes indicate conserved residues. The positions of the three aspartic acid (D) residues and QxxRW motifs are indicated by asterisks. Positions of the presumed membrane-spanning

helices are indicated by solid bars. Variable regions referred to in the text are also indicated (VR1 and VR2). Dots were introduced to optimise alignment.

Figure 4 shows toluidine stained sections of Arabidopsis vascular bundles from wild-type, *irx3*, and *irx3* plants transformed with cosmids L1, L10, L3 and L5.
Co, cortex; ph, phloem; xe, xylem elements.

Figure 5 illustrates cellulose measurements showing complementation of the *irx3* cellulose deficient phenotype using cosmid clones.
Cellulose content of stem sections from individual wild-type (WT) and *irx3* plants together with individual *irx3* plants transformed with cosmids (L1, L3, L4, L5, and L10) containing the *IRX3* gene. Details of the cosmids are provided in Figure 2.

Figure 6 shows RNA gel blots showing expression of the *IRX3* gene.
Blots containing RNA from developing stems and leaves from wild-type (wt) and *irx3* plants were probed with 75G11, COMT and rRNA.

Figure 7 illustrates a phylogenetic tree of bacterial and plant cellulose synthases and homologues. Alignment data were bootstrap sampled 100 times and used to construct the consensus tree shown. Numbers are bootstrap values and indicate the number of trees in which the sequences to the right of a bootstrap value clustered together. Shown to the right of *Csa*, *Csb* or *Csc* gene names are the GenBank accession numbers for each gene. Agrobacterium refers to *A. tumefacians*, Acetobacter for *A. xylinum*, and Aquifex for *A. aeolicus*.

Fig 8 A and B show transverse sections through the base of immature inflorescence stems of Arabidopsis plants transformed with the *IRX3* promoter-uidA construct. White boxes indicate the extent of the xylem and the black box the extent of the interfasicular region. co - cortex; ph - phloem; pi - pith.
C and D show whole root mounts of *IRX3-uidA* transgenic seedlings. Root hairs are seen radiating from the main root.

Flg 9 shows Gus staining of tobacco stems transformed with pp8GUS. Staining is localised to areas of developing xylem, such as the xylem of a developing side shoot (top), or on the inner side of the vascular cylinder where new primary xylem is forming (bottom).

EXAMPLES

Library Screening

Standard molecular techniques were carried out as described in Sambrook *et al.*, (1989). A Landsberg *erecta* library constructed in lambda FIX (Voytas *et al.*, 1990) was screened with a 1.4 kb SalI -XbaI fragment from expressed sequence tag (EST) clone 75G11, labelled nonradioactively with the Gene Images random prime labelling module (Amersham Life Science, Little Chalfont, Buckinghamshire, UK) probed and developed with the Gene Images CDP-Star detection module (Amersham Life Science) according to the manufacturer's instructions before visualisation of signal on BioMax MR1 film (Eastman Kodak, Rochester, New York). Two rounds of screening were carried out to identify hybridising clones.

Cosmids carrying *IRREGULAR XYLEM 3* (*IRX3*) were isolated from a Landsberg *erecta* library constructed in pBIC20 (Meyer *et al.*, 1994). Filters carrying 120,000 library clones were hybridised with a random primed digoxigenin-11-2'-deoxyuridine-5'-phosphate–labeled 200 bp polymerase chain reaction (PCR) fragment, amplified by using primers 75G11F and 75G11R (see Results), and developed, and the positive clones were detected colorimetrically as described by the kit manufacturer (Boehringer Mannheim, Germany). Two rounds of screening were carried out to identify cosmid clones harbouring *75G11* genomic DNA.

RNA Gel Blot Analysis

Total RNA was isolated from 6-week-old plants using an RNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany). After transfer of 5 μg electrophoresed RNA to Hybond N+ membranes (Amersham Life Science) they were probed with 75G11 (1.4 kb SalI / XbaI fragment), COMT (Arabidopsis Biological Resource Center, Columbus,OH, stock center clone 115N5, EcoRI / HindIII 1.5 kb fragment), or rRNA (O'Donnell *et al.*, 1998; 300 bp EcoRI fragment) probes labelled as given above and developed according to the manufacturer's instructions before visualisation as above.

PCR and Reverse Transcription PCR

PCR was carried out using Taq polymerase (Immunogen International, Sunderland, UK) according to manufacturer's recommendations in a PTC100 thermal cycler (MJ Research Inc,

Watertown, MA). Yeast artificial chromosomes (YAC) template DNA was isolated using an IGi Yeast Y1-3 kit (Immunogen International). Oligonucleotide primers were synthesised either by Gibco BRL Life Technologies UK Ltd. (Paisley, UK) or MWG Biotech UK Ltd. (Milton Keynes, UK). Primer sequences for polymerase chain reaction (PCR) of 75G11 from YAC clones are as follows: 75G11F, 5'-AAGGTGATAAGGAGCATTTGA-3' (SEQ ID NO. 5) and 75G11R 5'-TCCCCACTCAGTCTTGTCTT-3' (SEQ ID NO. 6). The PCR conditions were as follows. 94°C for 60 sec followed by 10 cycles of 94°C for 45sec, 65°C for 60sec (reducing by 0.5°C per cycle), and 72°C for 60 sec followed by 25 cycles at 94°C for 45 sec, at 55°C for 60 sec and 72°C for 60 sec followed by 5 min at 72°C.

For RT-PCR, first-strand cDNA was synthesised using 500 ng of mature stem total RNA in a reaction with a Ready To Go RT-PCR Bead (Pharmacia Biotech, Uppsala, Sweden) with 500 ng poly (dT) primer at 42°C for 60 min. Gene specific primers IRX3F (5'-CCTATGGAAGCTAGCGCCGGTCTT-3') (SEQ ID NO. 7) and IRX312 (5'-GTGTTTCTGTTGGCGTAACGA-3') (SEQ ID NO. 8) were added for the 5' end of the cDNA, and IRX3R (5'-GCTTCAGCAGTTGATGCCACACTT-3') (SEQ ID NO. 9) and IRX315 (5'-CGTTGAAAGTTGATTATCTCC-3') (SEQ ID NO. 10) were added for the 3' end. PCR conditions were as follows. 95°C for 5 min followed by 30 cycles at 94°C for 60sec, at 55°C for 60sec and 72°C for 2 min. RT-PCR products were gel purified before cloning into the vector pGEM-T Easy (Promega) for sequencing.

For PCR amplification from plant genomic DNA to ensure presence of the A-to-G nucleotide substitution, DNA was prepared from leaf tissue using a Phytopure plant DNA extraction kit (Scotlab, Lanarkshire, UK). Primers IRX33 (5'-TGCCTGCAACAACGCCAACAA-3') (SEQ ID NO. 11) and IRX317 (5'- TTGGGCACTTGGATCGGTTGA-3') (SEQ ID NO. 12) were used to amplify this fragment under the following conditions: 94°C for 60 sec followed by 30 cycles at 94°C for 60 sec, at 55°C for 60 sec and 72°C for 60 sec. Again, the products were gel purified and cloned into pGEM-T Easy for sequencing.

DNA Sequencing

Templates were generated by restriction fragment cloning or exonuclease III-generated deletions and primed with oligonucleotides annealing either to universal priming sites or gene specific regions. Sequencing primers were synthesised and HPLC or high purity salt free (HPSF) purified by MWG Biotech or PE Applied Biosystems. Plasmid templates were prepared using a Qiagen QIAprep Spin Miniprep Kit and sequenced automatically using ABI

PRISM Big Dye Terminators (PE Applied Biosystems, Foster City, CA). DNA sequence was analysed using the Genetics Computer Group suite of programs (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, Madison, WI) and programs available for use on the Internet.

Complementation of *irx*3

*irx*3 mutant plants were transformed by *Agrobacterium tumefaciens* (GV3101) carrying the appropriate Landsberg *erecta* binary cosmids according to Bent and Clough (1998). Primary transformants (T_1) were selected by plating sterilised T_1-seeds on Murashige-Skoog 0.8% agar plates containing 50 μg/ml kanamycin sulphate. After 3 weeks, the kanamycin-resistant plants were transplanted into pots containing a commercial soil/peat/perlite mixture. Stems from mature T_1-plants together with stems from same-aged Landsberg *erecta* wild-type and *irx*3 mutant plants were sectioned and stained with toluidine blue, and the cellulose content was then measured as described (Turner and Somerville, 1997).

Phylogenetic Analysis

Trees were built using PROTPARS, a maximum parsimony algorithm included in the PHYLIP version 3.5 software package (Felsenstein, 1993). Robustness of tree topology was estimated using 100 bootstrapped data sets (Felsenstein, 1985). These are generated by randomly sampling input alignment data until a new data set equivalent in size to the original is generated. Topologies observed in a large percentage of trees are believed to be robust (i.e., supported by multiple characters in the alignment data).

Sequences used for alignments were identified by BLAST searches of GenBank. Several expressed sequence tags (ESTs) with significant similarity to *IRX3* were excluded from our alignments. ESTs typically represent a small fraction of coding sequence, consequently we felt they did not posses enough useful (or reliable) sequence information to warrant inclusion in our data set.

Alignments were made using CLUSTALW (Thompson *et al.*, 1994). Initially, CLUSTALW failed to align the bacterial domain B residues (as defined by Saxena *et al.*, 1995) with the plant domain B residues. This was presumably due to the large insertion present within the plant domain B block. This problem was rectified by manually aligning the bacterial and plant domain B sequences by inserting gaps into the bacterial sequences. This alignment was refined with a second CLUSTALW alignment. Trees made with the initial and refined

alignment data sets were largely in agreement; both identified three deep branches separating the *CSA*, *CSB* and *CSC* gene families (data not shown). Not all residues of the alignment were used to build the tree shown in Figure 7; only sequence blocks conserved among the majority of sequences were used. These blocks include domains A, B, and other conserved regions visible in our alignment. With reference to the IRX3 sequence, the following sequence blocks were used: 320-359, 376-390, 497-512, 518-574, 581-606, 715-744, 750-781, 784-868, 883-907, 924-980.

Isolation of 3.2kb Promoter HindIII Fragment

The 7.5kb *Hin*dIII fragment isolated that carried the *IRX3* gene (Taylor *et al.*, 1999) was found to contain only 90bp of sequence upstream from the start codon. This made it necessary to isolate the 3.2 kb *Hin*dIII fragment that lay upstream of the 7.5kb *Hin*dIII fragment. DNA was isolated from cosmid L6 (Taylor *et al.*, 1999) and digested with *Hin*dIII. The 3.2kb fragment was then gel isolated before being ligated into pBluescript (Stratagene, La Jolla, CA, USA) before being completely sequenced on both strands. Oligonucleotide primers were designed in order to sequence across the junction with the 7.5kb *Hin*dIII fragment to ensure continuity.

Construction Of Promoter - GUS Fusions.

In order to determine the expression pattern of the *IRX3* promoter, it was decided to make a promoter GUS fusion. A number of vectors are available that allow the creation of a transcriptional fusion with the *uidA* gene that encodes β-glucuronidase, including the vector pCB1381Z (Jefferson, 1997). To clone a fragment of the *IRX3* promoter, PCR primers P1 (5'GC**GTCGAC**AGGGACGGCCGGAGATTAGCA3'(SEQ ID NO. 13), sequences complementary to *IRX3* promoter bases (1729-1749) underlined, *Sal*I site in bold) and P17 (5'GCAATC**CTCGAG**AGCCCGAG3'(SEQ ID NO. 14), entire sequence complementary to *IRX3* promoter (bases 1-14), *Xho*I site in bold) were used in a standard PCR reaction with cosmid L6 as template, and the resulting 1.75kb PCR product gel purified. This was then digested with *Xho*I and *Sal*I and ligated into pCB1381Z digested with *Sal*I, and the orientation of the insert confirmed with restriction digests, to create pP17GUS. This then consists of a 1749 bp *IRX3* promoter fragment controlling expression of the *uidA* gene. This plasmid was then transformed in Agrobacterium strain GV3101 in order that transgenic plants may be generated.

Transformation Of Arabidopsis.

Arabidopsis was transformed by vacuum infiltration (Bent and Clough 1998) with Agrobacterium carrying pP17GUS. Seeds from these plants were collected and transformants selected by plating on media containing 20mgl^{-1} Hygromycin. Transformed seedlings were then transfered to soil.

Analysis Of GUS Expression.

Staining of transgenic plants was carried out by immersing tissue in GUS histochemical buffer (Rodrigues-Pousada *et al.*, 1993. The Plant Cell 5:897-911.) before clearing in 80% ethanol and viewing. Whole seedlings were stained and mounted whole, and lengths of stem were stained before hand-cut sections were cut and mounted.

RESULTS

Identification of a Cellulose Synthase EST Linked to *irx3*

Because of the specific defect in secondary wall cellulose deposition in the *irx3* mutant, we tested the possibility that one of the *CSL* or *CELA* sequences present in the Arabidopsis database corresponded to the *irx3* locus. *irx3* maps to the middle of chromosome V and is close to the marker nga106 (Turner and Somerville 1997). In a cross between the *irx3* mutant and wild type, no recombinants were observed between *irx3* and nga106 in an analysis of 200 F_2 mutants (data not shown). Figure 1 shows that *irx3* is placed between markers nga151 and R89998. This region is represented by the seven CIC yeast artificial chromosome (YAC) clones CIC8E12, CIC9H7, CIC9F1, CIC6H3, CIC9E10, CIC11C4, and CIC6B10 (Creusot *et al.*, 1995, Schmidt *et al.*, 1997). Consequently the *irx*3 gene must be contained on one of these YACs.

Polymerase chain reaction (PCR) primer pairs were designed for each of the individual Arabidopsis *CELA* and *CSL* genes in GenBank, and each primer pair was tested to determine whether they amplified a fragment from the YAC clones spanning the region containing *irx3*. Only one of these primer pairs amplified a product, (75G11F and 75G11R), corresponding to the EST clone 75G11, amplifying a 200 bp fragment (data not shown). Analysis of the individual YACs in the region demonstrated that the 75G11 gene is contained on YACs CIC9H7, CIC9F1, and CIC6H3, but not on YACs CIC8E12, CIC11C4, CIC6B10, and CIC9E10 (Figure. 1). Based on the estimated relationship between physical and genetic map distance (Schmidt *et al.*, 1997), this information localised EST 75G11 to an approximately 150 kb region between markers nga106 and mi438 (Fig. 1). Because the *irx3* mutation also

maps between these two markers (results not presented), this information placed the EST 75G11 gene on a region of the chromosome, which was tightly linked to *irx3*.

Isolation of Genomic Clones Corresponding to EST 75G11

To obtain the full-length sequence of the gene corresponding to EST 75G11, the EST clone was used as a hybridisation probe to isolate genomic clones. A Landsberg *erecta* genomic library was screened and yielded two clones that were retained for characterisation. Figure 2A shows that one of these clones (pCS1) contains a HindIII fragment of 7.5 kb that was found to encode the entire coding sequence of the gene corresponding to EST 75G11. The nucleotide sequence of this fragment and the deduced amino acid sequence of the gene product has GenBank accession number AF091713. The cDNA sequence of the gene corresponding to EST 75G11 was determined by reverse transcription PCR (RT-PCR). To achieve this, primer pairs corresponding to the presumptive coding sequence, designed to amplify both the 3' and 5' halves of the gene, were used to amplify first strand cDNA. The fragments were cloned prior to sequencing. To negate the possible effects of incorporation of incorrect nucleotides by Taq polymerase, two independent clones isolated from individual RT-PCR reactions were sequenced and found to be identical (GenBank accession number AF088917).

Comparison of the cDNA and genomic sequences identified the presence of 11 introns and 12 exons in the genomic sequence. The cDNA sequence encodes a predicted protein of 1025 amino acids with a molecular mass of 116 kD. Figure 3 shows there is a high degree of sequence between the *75G11* gene product and several other cellulose synthase gene products, notably the Arabidopsis *RSW1* and *Ath-A* genes (Arioli *et al.*, 1998a) and the cotton *CELA1* gene (Pear et al., 1996). It is clear that there are significant regions of very high conservation. The only areas with no notable homology are in a region (VR2) that has been described previously as a plant hyper variable region (HVR, Pear et al., 1996) and a region close to the N terminus (VR1) (Figure. 3). In common with other cellulose synthase genes that have been identified (Pear et al., 1996; Arioli et al., 1998a), the *75G11* gene product contains a cysteine-rich region at its N terminus, which has been suggested to form a LIM-like Zinc finger motif which may be involved in protein-protein interactions (Delmer 1998). As expected, the *75G11* gene product also contains the four motifs that have been identified as being conserved in cellulose synthase genes. The first three of these are centred around aspartate residues, and the fourth consists of a QxxRW motif (where x represents any amino acid), which in this case, as in several other cases contains the sequence QVLRW (Figure. 3).

In common with cotton *CELA* and Arabidopsis *RSW1* (Pear *et al.*, 1996; Arioli *et al.*, 1998a), the *75G11* gene product shares a predicted transmembrane topology consisting of two transmembrane domains at the N terminus followed by a cytoplasmic central domain containing the four conserved motifs described. Six putative transmembrane segments at the C terminus follow this domain (Figure. 3).

Isolation of a Mutant Allele of *irx3*

To test the hypothesis that the *75G11* and *IRX3* genes are identical, the sequence of the *75G11* gene in the *irx3* mutant was determined. RT-PCR was used to isolate cDNA clones of the mutant allele. The cDNA was amplified in two halves, with two independent reactions carried out to control for the possibility of nucleotide misincorporation by Taq polymerase. Both clones showed a G-to-A nucleotide substitution, which resulted in the introduction of a stop codon in place of Trp-859. The region of genomic DNA containing this mutation was amplified by PCR and two independent products sequenced to confirm the presence of this mutation. Both products contained the G-to-A nucleotide substitution. This mutation causes premature termination of translation immediately after the second of the six carboxy terminal putative trans-membrane domains, and results in a protein lacking 168 C terminal amino acids. The identification of a mutation in the *75G11* gene in the *irx3* mutant strongly suggested that *75G11* is identical to *IRX3*.

Complementation of *irx3* with the Wild-Type Gene

To test whether the *irx3* mutation could be complemented with the wild-type gene, several cosmid clones containing the *75G11* gene were isolated and used to transform *irx3* plants. All of the cosmids contained a 7.5 kb HindIII fragment identified as carrying the coding region of the gene in its entirety (Figure. 2B). In addition, the clone contains 90 bp of sequence at the 5' end and 2603 bp at the 3' end of the gene.

Figures 4 and 5 show that cosmids L1, L4 and L10 (as well as L2, L6, and L8; data not shown) complemented the *irx3* mutation. Each of these contained the 7.5 kb HindIII fragment, an adjacent 3.2 kb HindIII fragment at the 5' end, and a 3.1 kb HindIII fragment at the 3' end of the *IRX3* gene (Figure. 2B). The 3.1 kb fragment carries no part of the *IRX3* coding region, and the nucleotide sequence of this fragment had no significant sequence similarity to any known genes as determined by BLASTX searches (Altschul *et al.*, 1990) against the Swiss Prot database. It can be seen from the transverse stem sections stained with

toluidine blue that in *irx3* plants, there is considerable collapse of the xylem vessels, whereas wild-type plants have clear, open xylem vessels (Figure 4). In plants transformed with cosmids L1 and L10, this collapse is not evident and these plants have xylem elements that are visually indistinguishable from those of the wild type. Cosmids L3 and L5, which did not carry the 3.2 kb fragment, failed to complement the mutation (Figure 4). In all plants transformed with L3, the xylem vessels exhibit the collapsed phenotype evident in the mutant, whereas in some of the plants transformed with cosmid L5 there was partial complementation of the mutant phenotype (Figure 4). This suggests that the requirement for the 3.2 kb 5' HindIII fragment is not absolute. The presence of this fragment is presumably necessary to direct correct expression of the gene. Because the 7.5 kb fragment carries only 90 nucleotides upstream of the coding sequence of the gene, the 3.2 kb fragment presumably contains the promoter required for normal correct expression of the gene. These promoter sequences are presumably found in the first 1.5 kb of this fragment, because the 5' end of this fragment appears to encode for part of a gene, which exhibits weak homology (BLASTX score 68, smallest sum probability $2e^{-33}$) to an APATELA2 domain-containing protein.

Measurements of the cellulose content of the primary transgenics (Figure 5) confirmed the results from qualitative analyses of xylem sections. Plants transformed with the cosmids L1, L4 and L10 contained cellulose levels that were indistinguishable from the wild type, whereas cosmid L3 had no effect on cellulose content. Thus, only cosmids that contained the 3.2 kb HindIII fragment effectively complemented the *irx3* mutation. Cosmids lacking this fragment (L3 and L5) did not complement, or only partially complemented the mutation.

Expression Patterns of the *IRX3* gene

RNA was isolated from leaves and from four discrete stem sections - the tip, upper middle part, lower middle part, and base of the stem of mature wild type and *irx3* plants. Figure 6 shows the results of probing this RNA with EST 75G11. In the wild type, there was an increase in the amount of *IRX3* mRNA as the stem matured (i.e., toward the base of the stem). There was no detectable transcript in leaves. These expression patterns correspond with secondary cell wall development. In comparison with the wild type, *IRX3* transcript levels were severely decreased in the *irx3* mutant, to approximately 10% of wild type levels in the most mature stem tissue (Figure 6). An identical blot probed with a gene encoding for caffeic acid *O*-methyltransferase (COMT), which is a component of the lignin biosynthesis pathway, showed that the *irx3* mutation had little effect on the expression of a typical gene in the lignin biosynthetic pathway (Figure 6). Minor differences in COMT transcript levels are

thought to be due to the difficulty in accurately staging the sections obtained from the different plants, because *irx3* plants have been shown to grow slightly more slowly than wild type (Turner and Somerville, 1997). Two possibilities exist as to the residual signal seen in *irx3* plants. It has been shown previously that the introduction of a premature stop codon into a transcript (as is the case with *irx3*) can lead to its degradation (Abler and Green 1996). Thus it would not be surprising if the message levels in *irx3* plants are reduced. It is not inconceivable that due to the close relationship between *CelA* like genes that there is some cross reaction with another member of the family, but the fact that the message level is decreased 90% in *irx3* shows that the large majority of the signal seen is derived from the correct message.

IRX3 is Part of a Large Family of Plant Cellulose Synthase Homologues

Analysis of current genomic sequence data indicates that Arabidopsis contains nine anonymous open reading frames with significant similarity to IRX3. Three other homologs have previously been described (Arioli *et al.*, 1998a). Thus, 13 Arabidopsis genes with significant similarity to *IRX3* are present in public databases. Because only about 30% of the Arabidopsis genome sequence is available, the size of this gene family is likely much larger. Proteins which share a common ancestor often share similar biochemical functions; understanding the evolutionary history of this gene family may help in future predictions of gene function.

To infer the evolutionary history of this gene family, a multiple alignment of plant and bacterial sequences similar to known cellulose synthases was constructed. The alignment data was bootstrap resampled and used to generate a maximum parsimony tree utilizing the PROTPARS algorithm (Felsenstein, 1993). The phylogenetic tree generated was rooted using a cellulose synthase homologue identified in the deeply branching prokaryote *Aquifex aeolicus* (Deckert *et al.*, 1998). Figure 7 shows the consensus tree generated by this analysis.

The phylogenetic tree reveals three deep branches, which divide the plant genes into three sub-families. These branches are supported by high bootstrap values and are unlikely to be spurious. Based on this data, we suggest that the higher plant family of sequences similar to IRX3 can be broken into three sub-families. To conform with Arabidopsis genetic nomenclature, we suggest these families be called CSA, CSB, and CSC (Figure 7). We intend for the CS prefix to indicate 'cellulose synthase homologue'.

The *CSA* gene family includes *RSW1*, *IRX3*, *CELA1* and *CELA2*. These genes are likely to be cellulose synthases based on either mutational analysis or expression data. Thus, the known plant cellulose synthase form a distinct sub-family within the gene family as a whole, and are not distributed throughout the family. The functions of the other branches remain to be determined. However, we believe they could function in the synthesis of one of many plant beta-linked polysaccharides (Cutler and Somerville, 1997).

After histochemical staining of *IRX3* promoter-*uidA* transgenic plants, staining was seen only in those cells that contain a thick secondary cell wall. Figs 8A and 8B show transverse sections from the base of the stem of *IRX3-uidA* transgenic plants, and it can be clearly seen that the GUS expression is localised to cells in the xylem (the clear cells being those cells which have undergone cell lysis and all that remains is a cell wall). There is also staining specific to cells in the interfascicular region, which show uniform staining in all cells, as there are no conducting elements in this region. It is clear that there is no GUS expression in the cortex, phloem or the pith, cell types which do not possess a heavily thickened secondary cell wall. Expression of GUS as directed by the *IRX3* promoter in roots is also very specific (Figs 8C and 8D), with expression being seen only in the central vascular cylinder but not the surrounding cortical and epidermal cells , nor in root hairs. It is clear that this expression is localised to cells in vascular cylinder where the xylem cells are found. Fig 8D also shows the expression of GUS in the development of new xylem cells in the formation of a lateral root.

Sequences.

Genomic sequence consisting of 7.5kb HindIII fragment: Genbank AF091713.

cDNA sequence : Genbank AF088917, SEQ ID NO. 1

Cellulose synthase encoded by 7.5kb HindIII fragment; SEQ ID NO. 2.

1750bp promoter sequence; SEQ ID NO. 3.

500 bp promoter sequence; SEQ ID NO. 4.

DISCUSSION

Stems of the *irx3* mutant contain approximately 20 - 30% of the amount of cellulose in mature stem tissue of wild type (Turner and Somerville, 1997). This results in an alteration of the physical properties of the stem and also leads to collapse of the xylem vessels due to an inability to withstand the negative pressure generated by water transport (Turner and Somerville, 1997).

Because of the specific defect in cellulose deposition in the mutant, we hypothesised that the *irx3* mutation may cause a defect in a subunit of cellulose synthase. To test this hypothesis, we first identified all of the EST and genomic sequences with sequence similarity to the Arabidopsis *CSL* genes and the *CELA* genes from cotton that were present in public databases. We then tested whether each of these sequences was present on the seven YAC clones that span the region of the genome where the *irx3* mutation had been genetically mapped. One EST (75G11) was found to be present on three of the relevant YACs and was, therefore, deemed a candidate clone for the *IRX3* gene. The observation that the *75G11* gene carries a nonsense mutation in the *irx3* background and complementation of the *irx3* mutation with cosmids carrying 75G11 confirmed the coidentity of *75G11* and *IRX3*.

IRX3 likely encodes a cellulose synthase catalytic subunit similar to other plant and bacterial cellulose synthase genes (Arioli *et al.*, 1998a; Pear *et al.*, 1996). It contains all of the conserved motifs that have been proposed to be essential for cellulose synthase activity (Arioli *et al.*, 1998a; Pear *et al.*, 1996). The expression pattern of the *IRX3* gene in Arabidopsis is consistent with the expectation for a gene involved in the synthesis of cellulose to be deposited in heavily thickened secondary cell walls. The increased levels of accumulation of *IRX3* mRNA in more mature stem tissue is consistent with the observation that the cellulose content increases towards the base of the stem. This expression pattern of the *IRX3* gene also correlates well with the *irx3*-conferred phenotype, which exhibits a large difference in cellulose content in mature stems compared to wild type, but little difference in leaves (Turner and Somerville 1997).

Further evidence that *IRX3* is not involved in cellulose synthesis in primary walls derives from observations that *IRX3* does not exhibit any of the radial swelling phenotype or other phenotypes characteristic of the *rsw1* mutant, despite the very severe nature of the *irx3* mutation, which suggests it is probably a null mutation. In addition, whilst *rsw1* mutants plants exhibit a decrease in crystalline cellulose there is an increase in non-crystalline β-1-4 linked glucose (Arioli *et al.*, 1988a). *irx3* plants apparently show no increase in this non-crystalline β-1-4 linked glucose, since despite the very large decrease in crystalline cellulose observed in *irx3*, no increase has been observed in the proportion of glucose in the non-crystalline (soluble in 2M sulphuric acid) cell wall fraction (Turner and Somerville 1997). Until the definitive confirmation that recombinant proteins produced from these genes actually have cellulose synthase activity, it is still possible that these genes may encode, for

example, a protein that primes rather than extends the cellulose chain. The work presented here, however, adds to the growing body of evidence (Pear *et al.*, 1996, Arioli *et al.*, 1998a) that these genes do in fact encode for the catalytic subunit of the higher plant cellulose synthase complex.

The relatively large number of cellulose synthase like (*CSL*) sequences from Arabidopsis that are present in public databases have raised questions as to the function of these sequences (Cutler and Somerville, 1997). The results presented here indicate that the function of at least some of the genes may be accounted for by cell-type specific gene expression. Similarly, in the *rsw*1 mutant, epidermal cells are misshapen (Arioli *et al.*, 1998a), and it is possible that only this cell type is affected. It has been suggested that of the ~40 cell types present in plants, almost all can be identified by unique features of their cell walls (Carpita and Vergara, 1998). In light of this, it may not be surprising that different cell types may utilise individual sets of genes for their cell wall synthesis.

The inferred phylogenetic relationship between the cellulose synthase genes aligned in Figure 3 and some genes that have been suggested to be more weakly related (Arioli *et al.*, 1998b) is shown in Figure 7. It is clear that *IRX3* belongs to a small subfamily of cellulose synthase genes, including *RSW1* and cotton *CELA1*, but shows distant relationships to a large number of other cellulose synthase-related genes. This supports the idea that only the CSA subfamily of genes is involved in cellulose synthesis, whereas the function of other cellulose-synthase related genes remains unknown (Arioli *et al.*, 1998a and b). It can be seen that *IRX3* is closely related to *Ath-B*, an Arabidopsis cDNA of unknown function isolated by screening a cDNA library with a portion of the *RSW1* transcript (Arioli *et al.*, 1998a) and to a gene, which we have provisionally named *CSA1*, that is evident in the currently available Arabidopsis genomic DNA sequence. *IRX*3 also appears to be more closely related to the *CELA*1 and *CELA2* genes from cotton (Pear *et al.*, 1996) than it does to the Arabidopsis *RSW1* gene (Arioli *et al.*, 1998a), based upon the results of PILEUP analysis (data not shown). Thus, it seems possible that *IRX3*, *CELA*, *CSA1* and *Ath-B* are all involved in secondary wall synthesis, whereas *RSW1* and *Ath-A* define the class of enzymes involved in primary wall synthesis.

Comparison of these sequences may make it possible to identify features that identify what type of cell wall is produced by a particular cellulose synthase. Do cellulose synthases involved in secondary cell wall synthesis contain some sequences, which allow them to form

rosette structures that cluster, to produce larger cellulose microfibrils? It is clear that there are two regions of variability between plant cellulose synthase genes. One of these lies close to the amino-terminal region that is predicted to be cytoplasmic and also contains a putative cysteine-rich LIM-like protein binding domain (Delmer, 1998). We speculate that this is a region of the protein involved in interactions with other proteins that may make up the enzyme complex found in the membrane and possibly with other regulatory proteins as well. It should be noted, however, that there is another region of variability that has been called an HVR (Pear *et al.*, 1996). This region lies between the second and third conserved motifs, and as such could be involved in the catalytic process itself. It is clear that there is still much to be learned about the synthesis of cellulose, with many important questions to be answered concerning the number of genes actually encoding cellulose synthases, and their possible differences in laying down cellulose. The catalytic mode of action of cellulose synthase is also an area in which advances need to be made to further our understanding. The cloning of *IRX3*, a gene involved in the synthesis of cellulose in secondary cell walls, will allow us to investigate some of these matters. For instance, it will be instructive to test whether *RSW1* or any of the other *CELA*-like genes will functionally complement the *irx3* mutation.

The mutation in the *irx3* mutant leads to the loss of the last 168 amino acids of the mature protein. This portion contains four membrane-spanning domains and several other features conserved in *RSW1* and *CELA1*. Its is very unlikely that such a gene would retain catalytic function and, therefore, the *irx3* mutation appears to be a null mutation. In support of this conclusion, electron microscopy of sections of stems from *irx3* plants show little if any cellulose in the secondary cell wall of xylem cells (Turner and Somerville, 1997). Nevertheless, under laboratory conditions, *irx3* plants can grow and produce relatively normal plants in the absence of a normal secondary cell wall. Thus, it should be possible to recover any mutation that inactivates the cellulose synthase specifically required for secondary wall synthesis. However, if the same genes are used for components of both the primary and secondary walls, it may not be possible to identify nonconditional mutations in these genes. In this respect, the characterisation of the *irx1* and *irx2* mutations (Turner and Somerville, 1997) may provide additional insights into the process of cellulose synthesis and deposition.

The identification of the *IRX3* gene was greatly facilitated by analysis of publicly available sequence data. In the near future, this sequencing initiative is likely to be an area of plant research that will revolutionise the way in which gene functions are assigned. The only other

report involving the cloning of a cellulose synthase gene from Arabidopsis involved a long chromosome walk to the gene (Arioli *et al.*, 1998a). The increasing number of ESTs that are easily mapped using PCR-based methodology and the completion of Arabidopsis genome sequencing, should soon supersede the need for such chromosome walks and will greatly accelerate the identification of genes responsible for mutations.

ADDITIONAL EXAMPLE

The pp17GUS construct (now termed pp8GUS) comprising a 1749 bp *IRX3* promoter fragment controlling expression of the *uidA* gene was used for the transformation of tobacco to show that the *IRX3* promoter can work in species other than Arabidopsis. Transformations were performed on tobacco leaves using Agrobacterium according to standard procedures. Staining of free hand sections was performed by incubating sections of developing stems from primary tobacco transformants in X-gluc as described previously for Arabidopsis. Presence of the reporter gene and hence pp8 promoter activity is indicated by the presence of a blue colour in those tissue in which the promoter is active as shown in Figure 9.

Further experiments were performed to hook up a lignin biosynthesis gene for work in co-suppression experiments in Arabidopsis to show that the cellulose synthase promoter can modulate the expression of a lignin biosynthesis gene.

Arabidopsis plants were stably transformed with the pp8 promoter in front of cDNA for the lignin biosynthesis gene courmaryl CoA reductase (CCR) using Agrobacterium, using stand techniques. The effect on cell wall properties was measured using an Instron universal testing machine exactly as described by Turner and Somerville 1997. A randomly selected sample of 14 T2 transformed plants gave an mean bending modulus (measure of rigidity) of 539 KPa and stress at yield (measure of cell wall strength) of 6.013 MPa. Comparable experiments for wild type plants give a bending modulus of 2028 MPa and stress at yield of 15.55 MPa.

These data indicated that the strength of the stem and its rigidity are greatly reduced by pp8CCR construct. Since properties are determined by the properties of the cell wall (Turner and Somerville 1997), this is the result of the pp8 promoter being active in cell synthesising secondary cell walls most likely by reducing lignin content.

REFERENCES

Abler, M.L., and Green, P.J. (1996). Control of mRNA stability in higher plants. Plant Mol. Biol. **32**, 63-78.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment sequence tool. J. Mol. Biol. **215**, 403-410.

Arioli, T., Peng, L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J., and Williamson, R.E. (1998a). Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science **279**, 717-720.

Arioli, T., Burn, J.E., Betzner, A.S., and Williamson, R.E. (1998b). Response: How many cellulose synthase-like gene products actually make cellulose? Trends Plant Sci. **3**, 165-166.

Baskin, T.I., Betzner, A.S., Hoggart, R., Cork, A., and Williamson, R.E. (1992). Root morphology mutants in *Arabidopsis thaliana*. Aust. J. Plant Phys. **19**, 427-437.

Bent, A.F., and Clough, S.J. (1998). *Agrobacterium* germ-line transformation: Transformation of *Arabidopsis* without tissue culture. In Plant Molecular Biology Manual (Second Edition), S.B. Gelvin and R.A. Schilperoot, eds (Kluwer Academic Pub., Dordrecht, The Netherlands). Section B7 pages 1-14.

Blaschek, W., Koehler, H., Semler, U., and Franz, G. (1982) Molecular weight distribution of cellulose in primary cell walls. Investigations with regenerating protoplasts, suspension cultured cells and mesophyll of tobacco. Planta **154**, 550-555.

Brown, R.M. (1996). The biosynthesis of cellulose. J. Macromol. Sci.- Pure and Appl. Chem. **A33**, 1345-1373.

Brown, R.M., Saxena, I.M., and Kudlicka, K. (1996). Cellulose biosynthesis in higher plants. Trends Plant Sci. **1**, 149-156.

Carpita, N., and Vergara, C. (1998). A recipe for cellulose. Science **279**, 672-673.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chaboute, M.E., Durr, A., Fleck, J., Gigot, C., Camilleri, C., Bellini, C., Caboche, M., and Bouchez, D. (1995) The CIC library: A large insect YAC library for genome mapping in *Arabidopsis thaliana*. Plant J. **8**, 763-770.

Cutler, S., and Somerville, C.R. (1997). Cellulose synthesis: Cloning in silico. Current Biology **7**, R108-R111.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392**, 353-358.

Delmer, D.P. (1998). A hot mutant for cellulose synthesis. Trends Plant Sci. **3**, 164-165.

Delmer, D.P., and Amor, Y. (1995). Cellulose biosynthesis. Plant Cell **7**, 987-1000.

Emons, A.M.C., and Mulder, B.M. (1998) The making of the architecture of the plant cell wall: How cells exploit geometry. Proc. Natl. Acad. Sci. USA **95**, 7215-7219.

Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**, 783-791.

Felsenstein J. (1993) PHYLIP (phylgeny inference package) version 3.5. Distributed by the author. Department of Genetics, University of Washington,Seattle.

Fry, S.C. (1988) The Growing Plant Cell Wall: Chemical and Metabolic Analysis. Longman, Harlow, U.K. pp. 333-340.

Giddings, T.H., Brower, D.L., and Staehelin, A. (1980). Plasma membrane of *Micrastrias denticulata* associated with the formation of cellulose microfibrils in primary and secondary cell walls. J. Cell Biol. **84**, 327-339.

Jefferson, RA. 1997. A comprehensive set of modular vectors for advanced manipulations and efficient transformation of plants. Rockefeller Foundation Meeting of the International Program on Rice Biotechnology, 15-19 Sep 1997, Malacca, Malaysia.

Lister, C., and Dean, C (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J. **4**, 745-750.

Matthysse, A.G., White, S., and Lightfoot, R. (1995). Genes required for cellulose synthesis in *Agrobacterium tumefaciens*. J. Bacteriol. **177**, 1069-1075.

Meyer, K., Leube, M.P., and Grill, E. (1994). A protein phosphates 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science **264**, 1452-1455.

Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C.R. (1994). Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol. **106**, 1241-1255.

O'Donnell, P.J., Truesdale, M.R., Calvert, C.M., Dorans, A., Roberts, M.R., and Bowles, D.J. (1998). A novel tomato gene that rapidly responds to wound- and pathogen-related signals. Plant J. **14**, 137-142.

Pear, J.P., Kawagoe,Y., Schreckengost, W.E., Delmer. D.P., and Stalker, D.M. (1996). Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Nat. Acad. Sci. USA. **93**, 12637-12642.

RodriguesPousada, R.A., Derycke, R., Dedonder, A., Vancaeneghem, W., Engler, G., Van Montagu, M., and Vanderstraeten, D. 1993. The Arabidopsis 1-aminocyclopropane-

1-carboxylate synthase gene 1 is expressed during early development. The Plant Cell **5**:897-911.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Saxena, I.M., Lin, F.C., and Brown, R.M. (1990). Cloning and sequencing of the catalytic subunit of cellulose synthase from *Acetobacter xylinum*. Plant Mol. Biol. **15**, 673-683.

Saxena, I.M., Brown, R.M., Fevre, M., Geremia, R.A., and Henrissat, B. (1995). Multidomain architecture of -glycosyl transferases: Implications for mechanism of action. J. Bacteriol. **177**, 1419-1424.

Schmidt, R., Love, K., West, J., Lenehan, Z., and Dean, C. (1997) Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. Plant J. **11**, 563-572.

Taylor, N.G., Scheible, W-R., Cutler, S., Somerville, C.R., and Turner, S.R. 1999. The *irregular xylem* Locus of Arabidopsis Encodes a Cellulose Synthase Required for Secondary Cell Wall Synthesis. The Plant Cell **11**:1-12.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research **22**, 4673-4680.

Turner, S.R., and Somerville, C.R. (1997). Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell **9**, 689-701.

Voytas, D.F., Konieczny, A., Cummings, M.P., and Ausubel, F.M. (1990). The structure, distribution and evolution of the TA1 retrotransposable element family of *Arabidopsis thaliana*. Genetics **126**, 713-721.

SEQUENCES

SEQ ID NO. 1

AAGCTTTTCACACATAAAAACCAAACTTATTCGTCTCTCATTGATCACCGTTTTGT
TCTCTCAAGATCGCTGCTAATCTCCGGCCGTCCCTATGGAAGCTAGCGCCGGTCTT
GTCGCCGGTTCTCATAACCGTAATGAACTAGTCGTCATTCACAACCATGAAGAGG
TTTTTCACATTTACTTTTTCTCATCACTTACCAAAGTTTTTTTTTACCAACGCTAGT
AAAATATTATTGCATTTTTTCGTTTTATTTGGTTACTATTTCTTAAAAATTCGTATG
TTTTGGGAATAAAAATATGATCATTTTTTTAAATCATCTTCTTATTATGAGACAAA
AATTTATAATCTGTATTCTGTAGTTGCAATAATGTTGTAGAAAATTCATATCTTTG
TTAGCAAACATAATAATTTTGTTGGTAATATTAAGTTGAGAAGTCAGGTTTAACC
ATTTTAATCGCTGTCATTTTTTTTATTATCTTTACTTCATTAGAATCTTTTTTTGTTG
ATTAGAAATTTCAGGTTTTATTTCGTCTTTAAGGAACTTAAACTTTTGTGTAATTA
TAACAGCCAAAGCCTCTGAAGAATCTAGATGGACAATTCTGTGAGATATGTGGAG
ATCAGATTGGTTTAACAGTAGAAGGAGACCTCTTCGTAGCTTGCAATGAGTGTGG
TTTTCCGGCGTGTAGACCTTGCTATGAGTACGAGAGAAGAGAAGGAACACAAAA
CTGTCCTCAGTGTAAGACTCGTTACAAGCGTCTCAGAGGTAAGTTATTTATTAATC
TCCCTCTGCTCTTGTGTTGTTCGACGAAATGCCTCTATGAAAATTTAAAAAGGCTG
TTCCTTTTTTTAGTTTGAACTTGGAGAGTAATGATCTGTTTTTTGGTTTCTGAAGGA
AGCCCAAGAGTGGAGGGAGATGAAGACGAAGAAGATATTGATGATATTGAGTAT
GAATTTAATATCGAACATGAACAAGATAAGCATAAGCATTCTGCTGAGGCTATGC
TTTATGGGAAAATGAGCTATGGAAGAGGTCCTGAGGATGATGAGAATGGGAGAT
TCCCACCTGTTATAGCTGGTGGTCATAGTGGAGAATTTCCAGTTGGAGGAGGTTA
TGGTAATGGAGAACATGGGCTTCATAAGCGTGTGCACCCATATCCATCATCTGAA
GCTGGTGAGTCTCATGGAAATGTTAACTTACATATAGATTTAAGAATGTCTCACA
GTGATGATTAGTTAGGGTCATGCATATCTCCATATGTGCAAATAACATAAGTATG
AGGCCTTCCAGCTTAATAGTAGATAGGGACATAGTTTCATAAACATGGACTTTGG
GTTCTATTACATTCTTTCTATGAAATTCATCAGCAGACCCTTTTTCTAACAATACTT
TCCTTTTTGTTTATGTGTGTAATTTTAATGTGGTAGGGAGTGAGGGAGGATGGCGG
GAAAGAATGGATGACTGGAAGCTCCAGCATGGAAATCTTGGGCCAGAACCAGAT
GATGATCCTGAGATGGGACTGTAATGCCTCCACAAACATTTATCTAAGACATCAG
TTTTGTATGATTTGGATTCATGCTTACAAAATTTTGGATTTGACTGGAAATGGCTG
TAGGATCGACGAGGCACGGCAGCCACTCTCGCGGAAAGTTCCCATTGCCTCAAGC
AAGATCAATCCATATCGGATGGTCATCGTTGCTAGGCTTGTGATTCTAGCAGTTTT
TCTGCGGTATAGGCTCTTGAATCCAGTGCATGATGCTCTGGGATTATGGCTGACCT
CTGTGATCTGTGAAATCTGGTTCGCTGTCTCTTGGATTCTTGATCAGTTCCCCAAG
TGGTTCCCTATTGAACGTGAGACCTATCTAGATCGGCTTTCCCTCAGGTAAAAT
CCACAGATTCTCAAGTAGAAGTCTTAAAATCTATGACGTTGGAGTTTGGATGTAA
ATATTTTTTGTTTATACAGGTACGAGAGAGAAGGTGAACCAAATATGCTTGCCCC
TGTAGATGTCTTTGTCAGTACGGTGGACCCATTGAAGGAGCCTCCCCTCGTCACAT
CCAACACTGTGCTGTCAATCTTGGCCATGGACTACCCAGTTGAGAAAATCTCCTG
CTATGTCTCTGACGACGGTGCTTCAATGCTTACATTCGAATCTCTCTCGGAAACTG
CTGAGTTTGCAAGAAAATGGGTTCCCTTCTGTAAGAAATTCTCCATAGAGCCACG
GGCACCGGAGATGTACTTCACGTTGAAAGTTGATTATCTCCAGGACAAAGTCCAC
CCAACATTTGTTAAGGAACGTCGAGCCATGAAGGTCAGTGTATATCACCTGATCT
AGTTATACCACCACCCATCTTTTCACTATAATCTAAACTTCATAAGTGAAATTGAC
ATATGACAGAGAGAATATGAGGAGTTCAAGGTCAGGATCAACGCTCAAGTGGCG
AAGGCCTCAAAGGTTCCTCTAGAAGGTTGGATCATGCAAGATGGAACACCGTGGC
CAGGGAACAACACCAAGGACCACCCCGGTATGATCCAAGTCTTCCTCGGCCACAG
CGGAGGATTTGATGTCGAAGGGCATGAGCTTCCTCGGCTTGTGTACGTGTCCCGT
GAGAAGCGTCCTGGTTTTCAACACCACAAGAAAGCTGGCGCCATGAATGCCCTGG

TAATTTTCTTGATCTGCCTTGGACCAAACAAGAAACTGATCTCCGTGTCTTGGACC
TAACCTGATACTTCTGTCAGGTTCGAGTGGCAGGCGTACTCACAAATGCTCCTTTC
ATGCTGAACTTGGACTGTGATCACTATGTAAACAACAGCAAGGCCGTGAGGGAA
GCAATGTGTTTTTTGATGGATCCTCAGATTGGAAAGAAGGTCTGCTATGTTCAGTT
CCCTCAAAGGTTTGATGGCATTGACACAAACGATCGTTACGCCAACAGAAACACA
GTCTTCTTTGATGTAAGACTCAATTCATATTTTTCCAACTTCTGGTCTAAATGAAA
ATGTACCCTCTTGCTTACACTCTTGTTTGCTACAGATCAATATGAAAGGTCTAGAT
GGAATCCAAGGTCCAGTTTACGTTGGTACTGGTTGTGTTTTCAAACGACAAGCTCT
GTATGGTTATGAACCACCAAAGGGTCCTAAACGTCCAAAGATGATAAGCTGTGGT
TGTTGTCCTTGCTTTGGGCGCCGGAGAAAGAATAAGAAATTTTCCAAGAATGACA
TGAATGGTGACGTAGCAGCCCTTGGAGGTAAATTATCCCAACAACCTTATAATAT
CAGTCCATTCTTGCAGTAGATTTCGTTTATGTTGGAATCTTGCGGATCTGATAGTG
TTTTTTGGCAGGAGCAGAAGGTGATAAGGAGCATTTGATGTCTGAAATGAACTTT
GAGAAAACATTTGGGCAATCATCCATCTTTGTAACCTCAACTTTGATGGAAGAAG
GTGGTGTTCCTCCGTCATCAAGTCCTGCAGTGCTCCTTAAAGAGGCAATCCATGTC
ATAAGCTGCGGTTATGAAGACAAGACTGAGTGGGGAACTGAGGTAATAATACTG
AATCGTAGAAATCACCTTCTTATTTGTGATTTAGTAGCTGGCCATTGTAAAAAACG
TTTGTGTATCTCGAAATTGCAGCTGGGTTGGATCTATGGCTCTATCACAGAGGATA
TTTTGACGGGATTCAAGATGCATTGCCGTGGATGGAGGTCTATTTACTGCATGCCT
AAGAGGCCTGCATTCAAAGGTTCAGCTCCTATTAATCTATCAGACAGGTTAAACC
AGGTTTTGCGTTGGGCACTTGGATCGGTTGAGATATTTTTCAGCCGGCACAGTCCT
CTCTGGTATGGCTACAAAGGAGGCAAACTCAAGTGGCTTGAGCGTTTTGCTTATG
CCAACACAACAATCTACCCCTTCACATCTATACCACTTCTTGCCTACTGTATCCTT
CCAGCCATCTGTCTCCTTACTGACAAATTCATCATGCCACCGGTTAGTAAAATTAT
CAGAGAAAAGCACTTAGAAGCTGCATCAAATGTGCTAACTATCTGTTTTCCCAAT
TTTTCTTTCAGATAAGCACATTTGCTAGTCTCTTCTTCATCTCACTGTTTATGTCGA
TCATTGTAACGGGAATCTTGGAATTGAGATGGAGCGGAGTTAGCATTGAGGAATG
GTGGAGAAACGAGCAATTCTGGGTCATTGGAGGAATCTCAGCTCATCTCTTTGCG
GTTGTCCAAGGTCTCCTCAAAATCTTAGCAGGCATTGACACAAACTTCACCGTCA
CATCAAAGGCAACAGATGATGATGACTTTGGAGAACTTTACGCATTCAAATGGAC
AACACTGCTGATCCCTCCAACAACTGTCTTAATCATAAACATTGTTGGCGTTGTTG
CAGGCATCTCAGATGCCATTAACAATGGATATCAGTCTTGGGGACCTCTATTTGG
TAAACTCTTCTTCTCCTTTTGGGTCATTGTTCATCTCTACCCATTCCTCAAAGGTCT
GATGGGTAGACAGAACAGAACACCAACCATTGTGGTGATTTGGTCAGTGTTATTG
GCATCTATCTTCTCTTTGCTTTGGGTAAGAATTGATCCTTTTGTGCTCAAGACCAA
AGGACCTGACACTTCCAAGTGTGGCATCAACTGCTGAAGCAAAATCTTTTTCGTC
TTCTGAAACTTTTTCTGTACTTTGTCGAGAGTGTTTTCACTTCTCTCTTTGGATTGC
ATAATTGGATTTTGTTTATTGTATATTAGCCAGTAAAACAGATGGATCTGGGTTTA
TTGTGAGCGAGTAATGCATTGTAAGAAAATTTGAACAAAACTACTTTTTAGTTAC
ACAGTAAAAGATTCACAGATACTTCCCTCAAGAGACATGTAGTTGCAGTGAAAAC
CCACAAATCTTGAATGCAAATTTTTAACAGAGAGCCTGAGACTTGTTCTTATATCG
TGGAGTTCACAAACAAAATAAAGAAACGACAACAAAAGCCTAAACACACGCACA
CATCTCATAACAACTTCTCAAGCTGAATATCATTTGATAACATTAAAAAACAAAT
TGAATTGCCTGAGTTTGTTTAGTTAAAGTCTCACCTCTTATTCTTTCGAAACACCA
AGGATCCACTGGGAGAGATGAGTCAAGTATGGCTCAGCCTGGGAGCCTTTGCTCA
AGACTGGTCCTAAGCTCTGCAACTCCGAGTTGAACTCCCTTTCGAAATTGCTGGTA
GAACACGTAACAGACATATTAGAACAATACCGTGATAATGAAGTTCAAAAATTCT
AACAGCAACGGCTTTTGTAGAGATAGTGTAGCCATTAGTTTGAAAAACAATTCCT
TGTTACTACTTACAATATGGACTCTGTAACGGTTGAGTCTCTTATCATTGTCTTCT
GTGGATGAGACTGGCTATTGATGTCGATAGACGATGATATCAACCACTGTGTGGC
TGCAGCATATTGCAGACACACTGATTTCAACTTCTCCATTTTCTGCACCATCA

GTAAATATGAGTGTAAAATATTGATAAACATTACACTATTATTAGTGCGTTCAGA
ATTTTTGCTCTTGTTGCACGAAAGAGAAGCACATCACACTGTTACATCTGTCCTGA
GAAAACATAACATTTAGTGGAACAGCGACAAACCTTAAGGACGTCAGGTAAAAG
AAGCAAGCATCCTCTCAGACATTTATCAAGAAAAAAGTCGTGATGTTGTATCACC
TGCAAGATGATGGCACATTGTTCAATGAAGAATGTTTTCCCTGCAAACCAAAACA
AAGCAGGTCTCTACACTTACCTCATCTACACTCCTTGTGGATTGTAGTCTATCGTG
CATTACATGCCAATTTGGTTCGAGAACCTGAAGAGACATCCCACATAAAGACAAG
CAAAAACAACATTAGAAAAACTTAGAGAGGCAATAGTTGCTAAAATAAACACCA
AATCATCTCATAAAGAATTATGTCTTTATATACTTCTGGTGTGATTTCTCTCCTCTG
TTCATTGATAGTGTACTTATTTCTCCAGAATCACGTTTTCAAGAATGTATTTCTCG
AAAAACACAGTTGCAAGAAGTGCCTACATTTGTAAGTATGGTTTTTACGACTACA
GAGATATCTGGATTCAATATACAGAGTGTATGTCAGCAATTCTAGAAGTTAGTTT
TCCAGAGACTTACACCGCAAAATCACAATGGAAAACAGGATCTATGAATAGACA
ATCAGAATCTTTAAAAAACAAAAAAATATTCAGATTCCATTCTATTAGCAAAAAA
AAATGTTTTAAAAAACAAACAAAACAAAACAAAAAGAACATTCAGACTTGCCTC
GAATGTAAGATAATGTAGAAGGCTACTGATGAATTTAAGCATGCTACGGCAGAG
AAGCGATGATCTAAGAATTGCTGTACCCTTAGAGTTCATAGACCGAATCCCCTGT
AGAATTTTCACAGCAGTTTGTTAGCGAAGCAAAATGATTAAAATGGTTTTAGAAG
AAACCAAAACTGAAACAAATTACAAGCTACACAATATTATATCAGTCCTTACTTG
ATGTATCTGCCAAGCACCACAAAGCTGGCGTTCCACATGTTTGCAATGAAAAAGA
AAGCGGAAAATTAACTGGTACTTTGACAATGCTTTCTTTGATATGACAATAG
ATAGTGGCCATTGAACCTGAAACACAAAGCCACAACCAACAATCAAACAATTGA
GTAGCCAATCGAATATGCATTTGTCGATTACTTATATAAGTATGAATGATGACTCT
AGAAATGCATTAGTAGTTATATATAACAACCATTAATAGCCATTAAGCTGTGTTA
TTACTTATATATATCTCGACGAACAAAATAGTTCGAATCCAGATACAAACAGGTA
GTTGCAGACAAGGCAAGAAAGGAGCAGACCTTGTAGCTCAAGGAAAATGTCTCT
AGACCAGTGATGCTCATTGGGTCCTCAATACTATTGCTATCAGTGTCCTTGTGCAT
TCCCAAAGTCGTAAGCAATGAAGCTCGATCCTAAGATATGAATACAAAGTGAGA
ACATATAATGTACAAACTCTGTAAATTCATCACATCCAGGAAAAGACAAATTAAC
TGCTCACCACACAACAAGTCAAGTCTTCGTGGCGAGGATCTGCTGCAGCCGCTGT
GGTACGTAGAGCAAGATCTAGCAAAGACTAAATAATACACATCCATTTAGATTCC
AGCAAAGAAGTAGAGGCCATAACGATTTAATAATGGTAAGAAGCTT

SEQ ID NO. 3.

CTCGAGAGCCCGAGTCTACCTATTGGTAGTTCTGCGAAACGTCTCAAGGACGTTA
ACAATCCGGTTCCAGCTATGATGATTAGTAATAACGTTTCAGAGAGTGCAAATAA
TGTTAGCGGTTGGCAAAACACTGCGTTTCAGCATCAGGGAATGGATTTGAGCTTA
TTGCAGCAACAGCAGGAGAGGTACGTTGGTTATTACAATGGAGGAAACTTGTCTA
CCGAGAGTACTAGGGTTTGTTTCAAACAAGAGGAGGAACAACAACACTTCTTGAG
AAACTCGCCGAGTCACATGACTAATGTTGATCATCATAGCTCTACCTCTGATGATT
CTGTTACCGTTTGTGGAAATGTTGTTAGTTATGGTGGTTATCAAGGATTCGCAATC
CCTGTTGGAACATCGGTTAATTACGATCCCTTTACTGCTGCTGAGATTGCTTACAA
CGCAAGAAATCATTATTACTATGCTCAGCATCAGCAACAACAGCAGATTCAGCAG
TCGCCGGGAGGAGATTTTCCGGTGGCAATTTCGAATAACCATAGCTCTAACATGT
ACTTTCACGGGGAAGGTGGTGGAGAAGGGGCTCCAACGTTTTCAGTTTGGAACGA
CACTTAGAAAAATAAGTAAAAGATCTTTTAGTTGTTTGCTTTGTATGTTGCGAACA
GTTTGATTCTGTTTTTCTTTTTCCTTTTTTTGGGTAATTTTCTTATAACTTTTTTCAT
AGTTTCGATTATTTGGATAAAATTTTCAGATTGAGGATCATTTTATTTATTTATTA
GTGTAGTCTAATTTAGTTGTATAACTATAAAATTGTTGTTTGTTTCCGAATCATAA
GTTTTTTTTTTTTTTGGTTTTGTATTGATAGGTGCAAGAGACTCAAAATTCTGGTTT
CGATGTTAACAGAATTCAAATAGCTGCCCACTTGATTCGATTTGTTTTGTATTTGG

AAACAACCATGGCTGGTCAAGGCCCAGCCCGTTGTGCTTCTGAACCTGCCTAGTC
CCATGGACTAGATCTTTATCCGCAGACTCCAAAAGAAAAAGGATTGGAGCAGAG
GAATTGTCATGGAAACAGAATGAACAAGAAAGGGTGAAGAAGATCAAAGGCATA
TATGATCTTTACATTCTCTTTAGCTTATGTATGCAGAAAATTCACCTAATTAAGGA
CAGGGAACGTAACTTGGCTTGCACTCCTCTCACCAAACCTTACCCCCTAACTAATT
TTAATTCAAAATTACTAGTATTTTGGCGGATCACTTTATATAATAAGATACCAGAT
TTATTATATTTACGAATTATCAGCATGCATATACTGTATATAGTTTTTTTTTTGTTA
AAGGGTAAAATAATAGGATCCTTTTGAATAAAATGAACATATATAATTAGTATAA
TGAAAACAGAAGGAAATGAGATTAGGACAGTAAGTAAAATGAGAGAGACCTGCA
AAGGATAAAAAAAGAGAAGCTTAAGGAAACCGCGACGATGAAAGAAAGACATGT
CATCAGCTGATGGATGTGAGTGATGAGTTTGTTGCAGTTGTGTAGAAATTTTTACT
AAAACAGTTGTTTTTACAAAAAAGAAATAATATAAAACGAAAGCTTAGCTTGAA
GGCAATGGAGACTCTACAACAAACTATGTACCATACAGAGAGAGAAACTAAAAG
CTTTTCACACATAAAAACCAAACTTATTCGTCTCTCATTGATCACCGTTTTGTTCTC
TCAAGATCGCTGCTAATCTCCGGCCGTCCCT

SEQ ID NO. 4

GATCACTTTATATAATAAGATACCAGATTTATTATATTTACGAATTATCAGCATGC
ATATACTGTATATAGTTTTTTTTTTGTTAAAGGGTAAAATAATAGGATCCTTTTGA
ATAAAATGAACATATATAATTAGTATAATGAAAACAGAAGGAAATGAGATTAGG
ACAGTAAGTAAAATGAGAGAGACCTGCAAAGGATAAAAAAGAGAAGCTTAAGG
AAACCGCGACGATGAAAGAAAGACATGTCATCAGCTGATGGATGTGAGTGATGA
GTTTGTTGCAGTTGTGTAGAAATTTTTACTAAAACAGTTGTTTTTACAAAAAAGAA
ATAATATAAAACGAAAGCTTAGCTTGAAGGCAATGGAGACTCTACAACAAACTA
TGTACCATACAGAGAGAGAAACTAAAAGCTTTTCACACATAAAAACCAAACTTAT
TCGTCTCTCATTGATCACCGTTTTGTTCTCTCAAGATCGCTGCTAATCTCCGGCCGT
CCCT

CLAIMS:

1 An isolated nucleic acid molecule comprising a cellulose synthase gene specifically expressed during deposition of secondary cell walls in lignin containing cells.

2 An isolated nucleic acid molecule comprising a cellulose synthase gene specifically expressed during deposition of secondary cell walls in Arabidopsis.

3 An isolated nucleic acid molecule according to claim 1 or claim 2 comprising the sequence shown as SEQ ID No. 1.

4 An isolated nucleic acid molecule according to claim 1 or claim 2 comprising the complement of the sequence shown as SEQ ID No. 1.

5 An isolated nucleic acid molecule according to claim 1 or claim 2 comprising the reverse complement of the sequence shown as SEQ ID No. 1.

6 An isolated nucleic acid molecule according to claim 1 or claim 2 comprising the reverse of the sequence shown as SEQ ID No. 1.

7 An isolated nucleic acid molecule comprising a sequence having at least 80 % sequence identity with the nucleic acid molecule sequences of any one of claims 3 to 6.

8 An isolated nucleic acid molecule containing a promoter of an isolated cellulose synthase gene specifically expressed during deposition of secondary cell walls in lignin containing cells.

9 An isolated nucleic acid molecule containing a promoter of an isolated cellulose synthase gene specifically expressed during deposition of secondary cell walls in Arabidopsis.

10 A promoter according to claim 8 or claim 9 comprising the sequence shown as SEQ ID No. 3.

11 A promoter according to claim 8 or claim 9 comprising the complement of the sequence shown as SEQ ID No. 3.

12 A promoter according to claim 8 or claim 9 comprising the reverse complement of the sequence shown as SEQ ID No. 3.

13 A promoter according to claim 8 or claim 9 comprising the reverse of the sequence shown as SEQ ID No. 3.

14 A promoter according to claim 8 or claim 9 comprising the sequence shown as SEQ ID No. 4.

15 A promoter according to claim 8 or claim 9 comprising the complement of the sequence shown as SEQ ID No. 4.

16 A promoter according to claim 8 or claim 9 comprising the reverse complement of the sequence shown as SEQ ID No. 4.

17 A promoter according to claim 8 or claim 9 comprising the reverse of the sequence shown as SEQ ID No. 4.

18 A promoter comprising a sequence having at least 60 % sequence identity with the nucleic acid molecule sequences of any one of claims 10 to 17.

19 A nucleic acid construct suitable for transforming a plant cell, the construct comprising, in the 5'-3' direction:

(a) a cellulose synthase promoter according to any one of claims 8 to 18, and

(b) a nucleotide sequence of an exogenous gene;

the construct being arranged such that expression of the exogenous gene is under the control of the promoter.

20. A nucleic acid construct according to claim 19 in which the nucleotide sequence is in a sense orientation.

21. A nucleic acid construct according to claim 19 in which the nucleotide sequence is in an anti-sense orientation.

22. A nucleic acid construct according to any one of claims 19 to 21 in which the nucleotide sequence codes for an enzyme involved in synthesis of plant cell wall components.

23. A nucleic acid construct according to claim 22 in which the enzyme is involved in cell wall polysaccharide biosynthesis.

24. A nucleic acid construct according to claim 22 in which the enzyme is involved in cell wall protein biosynthesis.

25. A nucleic acid construct according to claim 22 in which the enzyme is involved in cellulose biosynthesis.

26. A nucleic acid construct according to claim 22 in which the enzyme is involved in lignin biosynthesis.

27. A nucleic acid construct according to claim 26 in which the nucleotide sequence encoding the enzyme involved in lignin biosynthesis is in an antisense orientation.

28. A transgenic plant cell transformed with a nucleic acid construct according to any one of claims 19 to 27.

29. A plant comprising a transgenic plant cell according to claim 28, or fruit or seeds thereof.

30. A plant according to claim 29 wherein the plant is a woody plant.

31. A plant according to claim 29 selected from the group consisting of alfalfa, rice, maize, oil seed rape, forage grasses, eucalyptus, pine, spruce, poplar, Arabidopsis and tobacco species.

32. A method for altering the cell wall of a plant by altering the activity of an enzyme involved in synthesis of plant cell wall components, the method comprising stably incorporating into the genome of the plant a nucleic acid construct according to any one of claims 19 to 27.

33. A method for producing a plant having altered lignin structure comprising

(a) stably transforming a plant with a nucleic acid construct according to claim 26 or claim 27 to produce a transgenic cell; and

(b) cultivating the transgenic cell under conditions suitable to produce a mature plant.

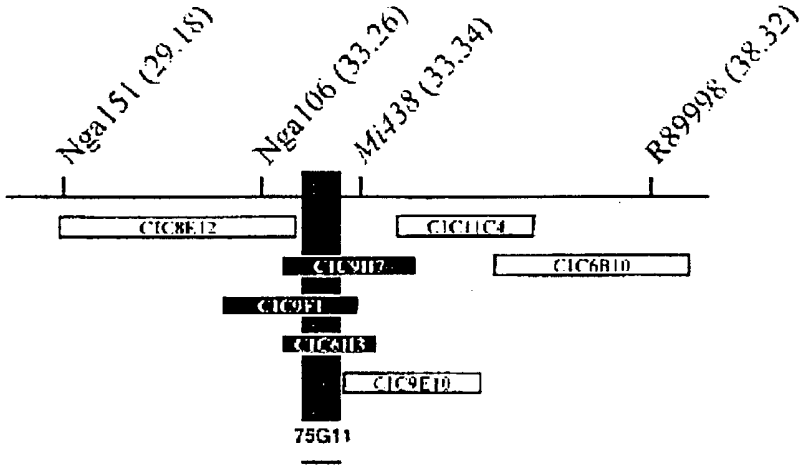

Nga151 (29.18)
Nga106 (33.26)
Mi438 (33.34)
R89998 (38.32)
C1C8E12
C1C11C4
C1C6B10
C1C9E10
75G11

FIG. 1

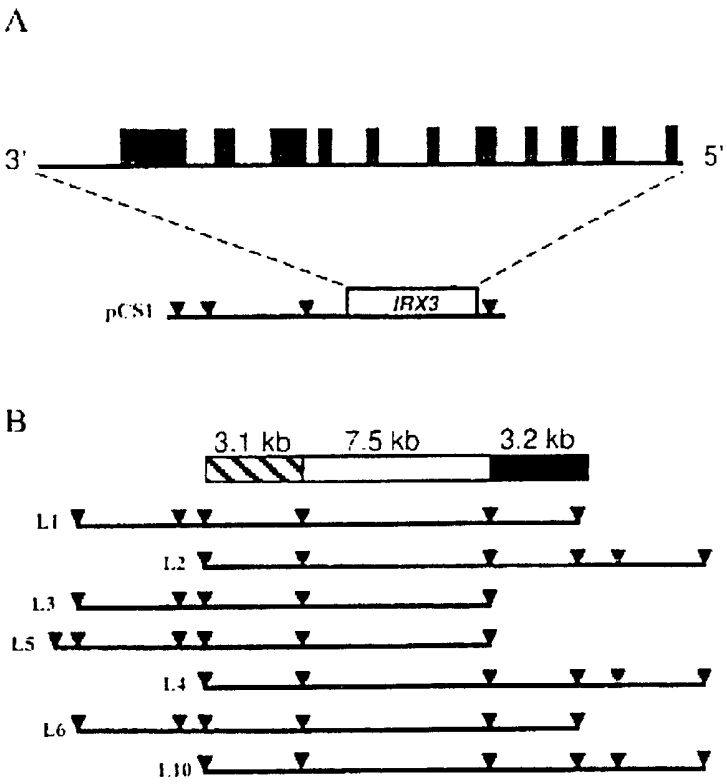


FIG. 2

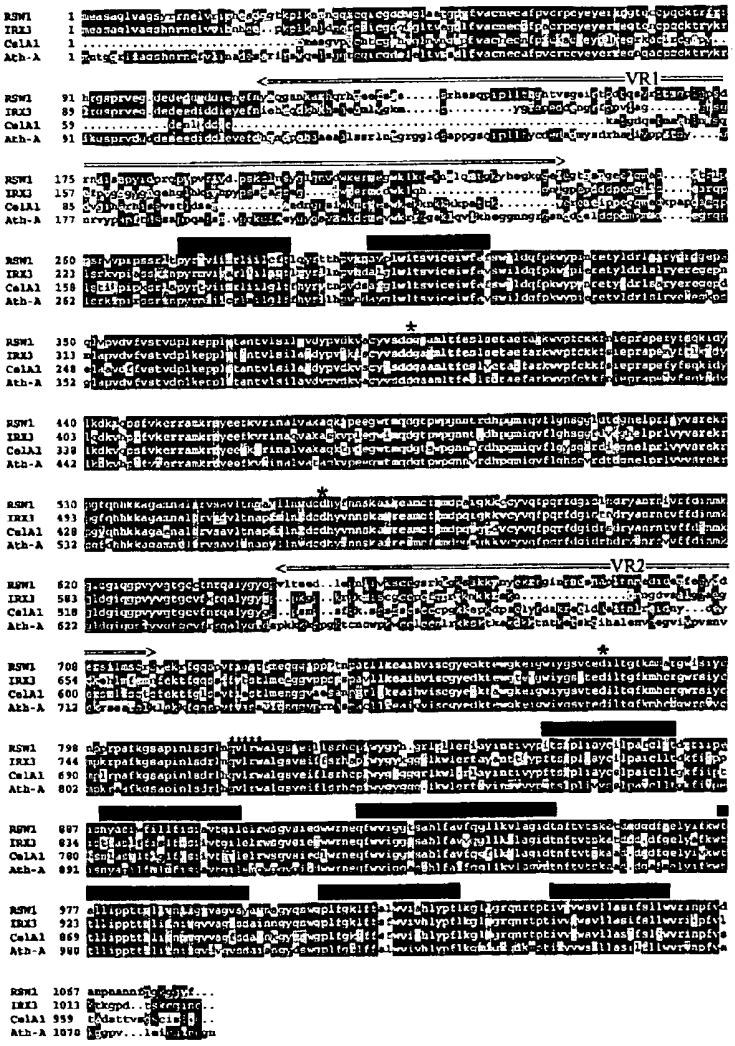


FIG. 3

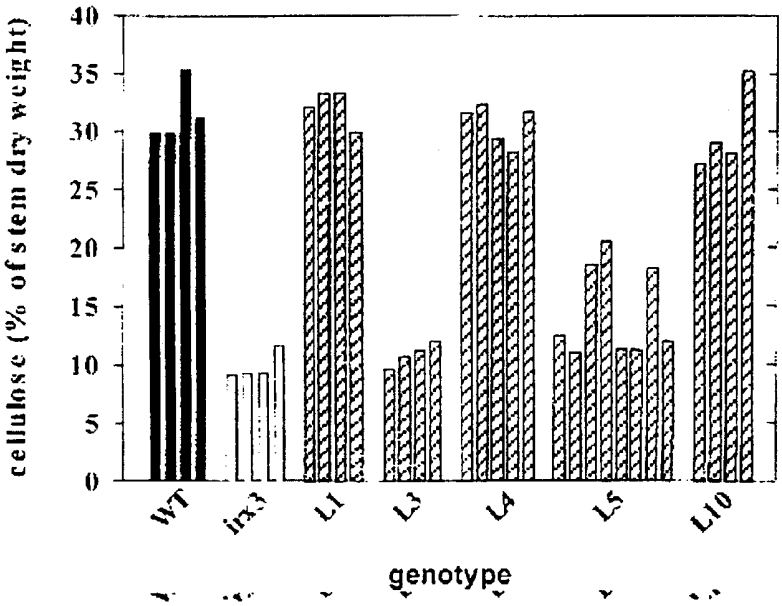

cellulose (% of stem dry weight)
40
35
30
25
20
15
10
5
0
WT
irx3
L1
L3
L4
L5
L10
genotype

FIG. 4

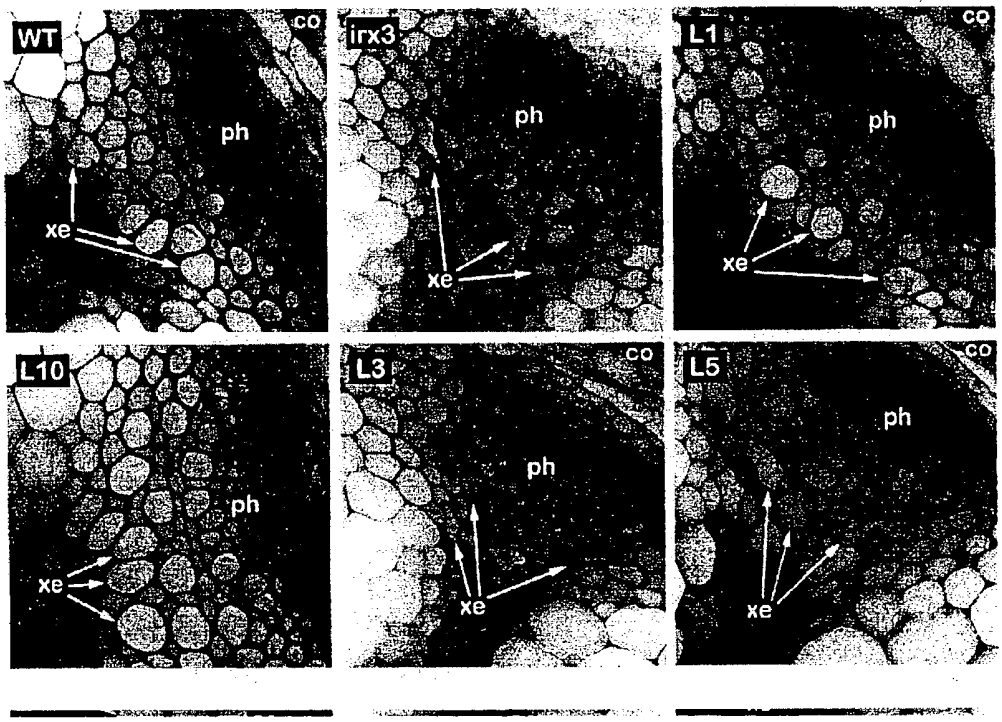


FIG.5

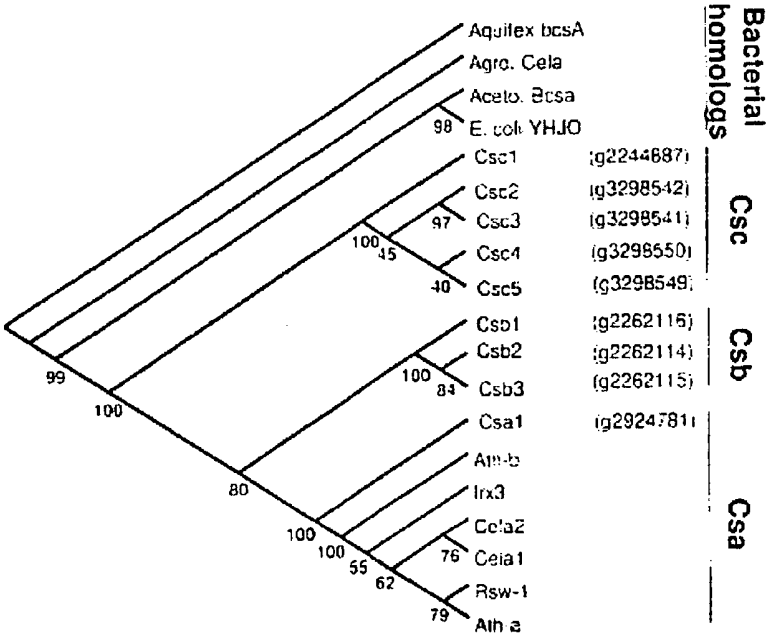

Aquifex bcsA
Agro. Cela
Aceto. Bcsa
E. coli YHJO
Csc1 (g2244687)
Csc2 (g3298542)
Csc3 (g3298541)
Csc4 (g3298550)
Csc5 (g3298549)
Csb1 (g2262116)
Csb2 (g2262114)
Csb3 (g2262115)
Csa1 (g2924781)
Alr-b
Irx3
Cela2
Cela1
Rsw-1
Alr a
98
99
100
97
100
45
40
100
84
80
100
100
55
62
76
79
Bacterial homologs
Csc
Csb
Csa

FIG. 6

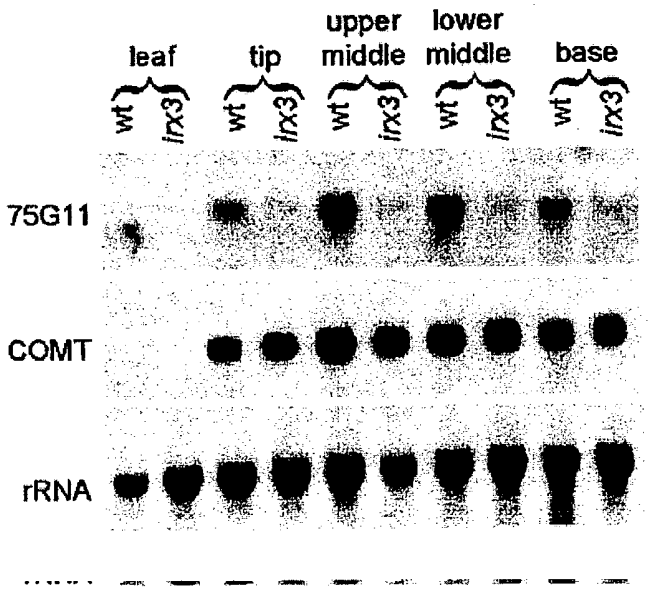

leaf
tip
upper middle
lower middle
base
wt
irx3
75G11
COMT
rRNA

FIG. 7

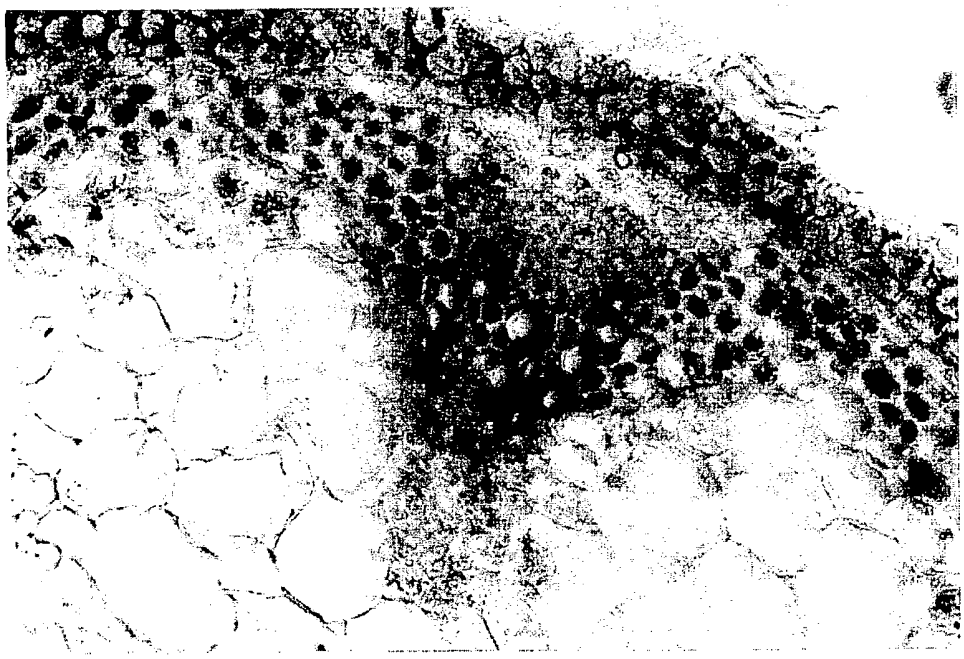


FIG.8a

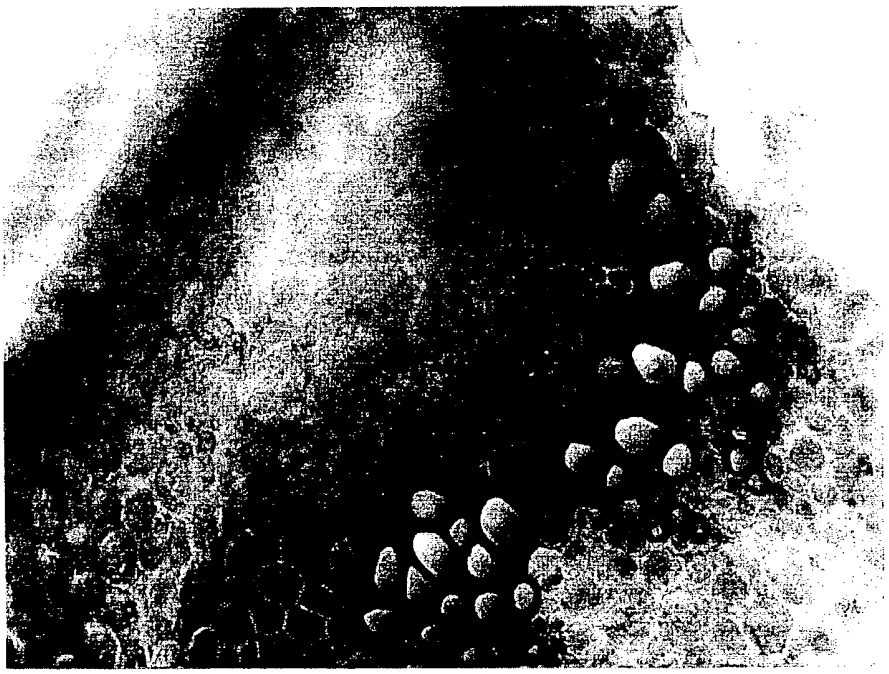


FIG.8b

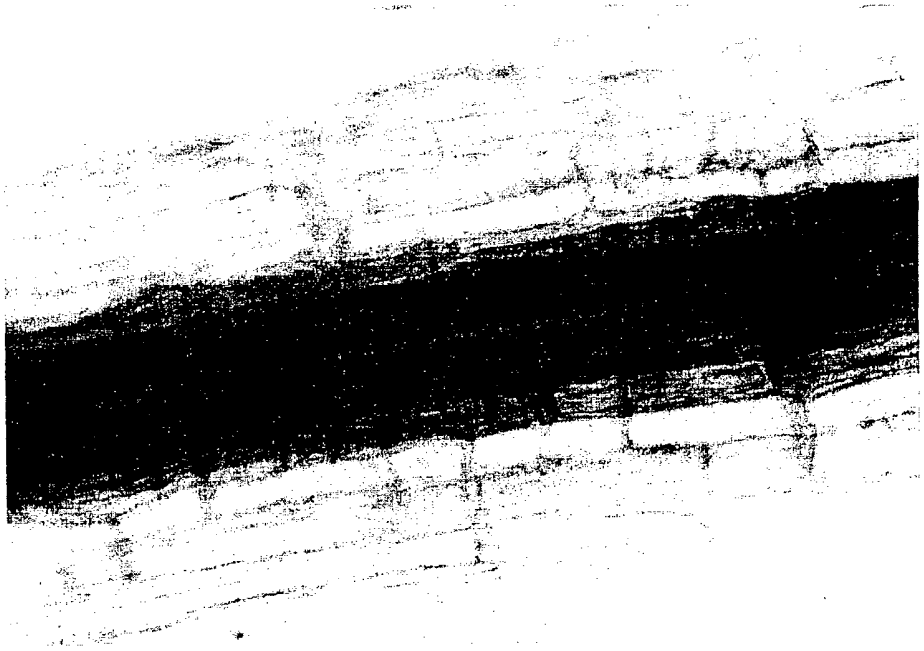


FIG. 8c

FIG. 8d

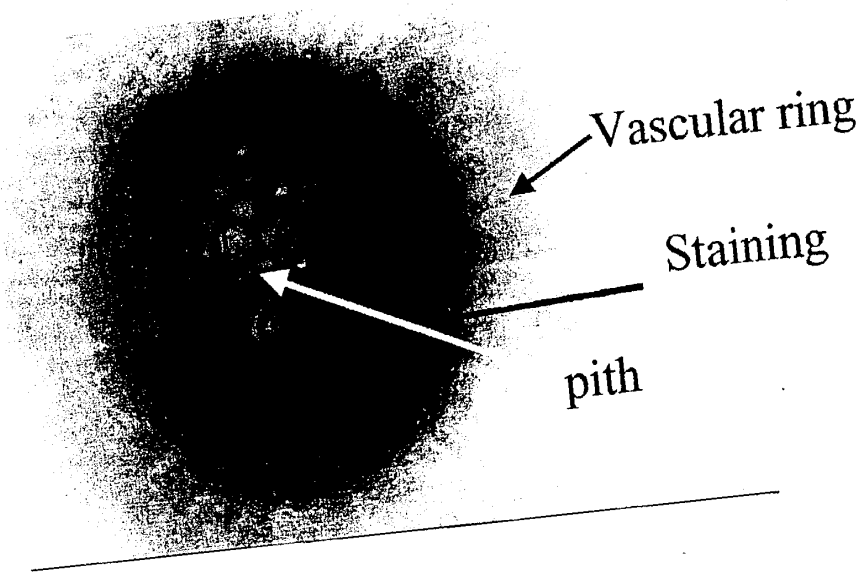


FIG.9